JBiSE

Fitting evolutionary process of matrix protein 2 family from influenza A virus using analytical solution of differential equation

Shao-Min Yan[1], Zhen-Chong Li[1], Guang Wu[2*]

[1]National Engineering Research Center for Non-food Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, China;
[2]Computational Mutation Project, DreamSciTech Consulting, 301, Building 12, Nanyou A-zone, Jiannan Road, Shenzhen, Guangdong, China.
Email: hongguanglishibahao@yahoo.com

ABSTRACT

The evolution of protein family is a process along the time course, thus any mathematical methods that can describe a process over time could be possible to describe an evolutionary process. In our previously concept-initiated study, we attempted to use the differential equation to describe the evolution of hemagglutinins from influenza A viruses, and to discuss various issues related to the building of differential equation. In this study, we attempted not only to use the differential equation to describe the evolution of matrix protein 2 family from influenza A virus, but also to use the analytical solution to fit its evolutionary process. The results showed that the fitting was possible and workable. The fitted model parameters provided a way to further determine the evolutionary dynamics and kinetics, a way to more precisely predict the time of occurrence of mutation, and a way to figure out the interaction between protein family and its environment.

Keywords: Amino-Acid Pair Predictability; Differential Equation; Evolution; Fitting; Influenza A Virus; Matrix Protein 2

1. INTRODUCTION

Very recently, we explored the possibility to use a differential equation to describe the evolution of hemagglutinin proteins from influenza A virus [1].

However, there are ten types of proteins from influenza A viruses, it is necessary to explore whether or not this differential description can be applied to other proteins from influenza A viruses. Also, it is intriguing to use the analytic solution of differential equation to fit the evolution of proteins from influenza A viruses.

This is so, because the mathematical modeling is generally the ending point of empiric experiments, and more importantly the mathematical modeling can provide us the tool for predicting the future of evolution.

As the evolution is a process along the time course, we at first needed to represent an evolutionary subject along the time course, and then we could consider how to apply the mathematical modeling to this process.

Now we are particularly interested in the evolution of proteins. However, a protein generally is a sequence of letters, which represent amino acids. Thus we need a method to represent a protein family along the time course before modeling [2-5].

In general, the evolution is a process of exchanging substances between a living subject and its environment. In this context, the differential equation is quite suitable because we defined $\frac{dy}{dt} = input - output$ for exchanging substance between a protein family and its environment along the time course [1].

As we know that the evolution of proteins goes through mutations, which bring in new mutating amino acids and take out mutated amino acids. This again requires the conversion of amino acids into numbers to represent the exchange [2-5].

Among ten types of proteins from influenza A virus, the matrix protein 2 (M2) is important because it constructs a proton channel in the virion and it is essential for infection [6]. Thus, the M2 protein was the target for anti-influenza drugs, and the M2 ion channel blockers was approved to treat influenza virus infections [7,8], but their use is limited by high frequencies of the resistance among currently circulating strains [9,10]. Also, a vaccine was designed basing on the conserved ectodomain of M2 protein, which could match multiple influenza virus strains including multiple subtypes [11].

In this study, our effort was made to apply the differential description to the evolution of M2 protein family from influenza A virus, and to use the analytical solution of differential equations to fit the evolutionary process of

Scientific Research

M2 proteins.

2. MATERIALS AND METHODS

2.1. Data

5926 full-length M2 proteins of influenza A virus sampled from 1959 to 2008 were obtained from the influenza virus resources [12]. After excluded identical sequences, 1084 M2 proteins were actually used in this study.

2.2. Conversion of Proteins into Numbers

For two purposes, we needed to convert M2 proteins into numbers: 1) we needed a single number for a single M2 protein so that we could present the evolution of M2 protein family over time, and 2) we needed a number to present mutation, which resulted in numerical exchange between M2 protein and its environment. We used the amino-acid pair predictability to do this job [2-5].

For example, an M2 protein (accession number ABF-01755) from an avian influenza virus, strain A/chicken/Magetan/BBVW/2005(H5N1), had 97 amino acids. The first and second amino acids could be counted as an amino-acid pair, the second and third as another amino-acid pair, the third and fourth, until the 96th and 97th, thus there were 96 amino-acid pairs.

This M2 protein had 10 glutamic acids (E) and 11 leucines (L): if the permutation could predict the appearance of amino-acid pair EL, it would appear once ($10/97 \times 11/96 \times 96 = 1.13$); actually it did appear once, so the pair EL was predictable. By contrast, this M2 protein had 7 isoleucines (I): if the permutation could predict the appearance of amino-acid pair IL, it would appear once ($7/97 \times 11/96 \times 96 = 0.79$); however, it appeared three times in realty, so the pair IL was unpredictable.

In this way, all amino-acid pairs in ABF01755 M2 protein were classified as predictable and unpredictable, which were 17.71% and 82.29%.

Taking another M2 protein (accession number ABF-01771) as example, this M2 protein had only one amino acid different from ABF01755 M2 protein at position 65. However, its predictable and unpredictable portions were 20.83% and 79.17%. Thus, the amino-acid pair predictability distinguished the difference between M2 proteins in numbers rather than in letters that represented amino acids in proteins.

Based on the above computation, the difference in predictable portion between ABF01755 and ABF01771 M2 proteins was −3.12% (17.71%−20.83%), which was regarded as the exchange between M2 protein and its environment.

2.3. Differential Equation

If ABF01755 and ABF01771 M2 proteins would have a direct relation due to a single mutation, the difference between them was $\frac{dy}{dt} = input - output$, where y was the difference in predicable portion, t was the time required for mutation, *input* was the predictable portion brought in by mutating amino acid, *output* was the predictable portion taken away by mutated amino acid.

Unfortunately, we had no way to know if ABF01755 and ABF01771 M2 proteins had a direct mutation relationship although both were sampled in 2005.

As the predictable portion was determined using the permutation based on random principle, thus this exchange was in fact the exchange of randomness between M2 proteins and their environment, more accurately was the exchange of entropy between M2 proteins and their environment [1].

2.4. Statistics

The Student's *t*-test and Mann-Whitney *U*-test were used to compare the difference between uphill and downhill half-life, and $P<0.05$ was considered statistically significant. The SigmaPlot for Windows was used for fitting [13].

3. RESULTS AND DISCUSSION

Ideally, we would hope to have a direct mutation relationship for all 1084 M2 proteins from 1959 to 2008 involved in this study, because then we would have $\frac{dy}{dt} = input - output$ for each mutation relationship between any related two M2 proteins.

Although it was impossible to find out such a relationship among all of these M2 proteins sampled everywhere in the world, the real mutation relationship worked in this way no matter if we had sampled them or not. Thus, we had a system of differential equations for all M2 proteins.

On the other hand, our computations on predictable portions of 1084 M2 proteins provided us with the most update evolutionary process in **Figure 1**, which was read as follows. For example, the solid curve in the top panel presented the evolution of 1084 M2 proteins from 1959 to 2008, and each point was the mean value of predictable portions of all M2 proteins in given year with its standard deviation (vertically grey line). The similar reading was applied to other panels.

So the fluctuating solid curves in **Figure 1** presented the evolutionary process over time. If we could use the differential equation to describe these solid curves, it would mean that we were able to model the evolutionary process of M2 proteins.

3.1. Possibly Analytical Solution

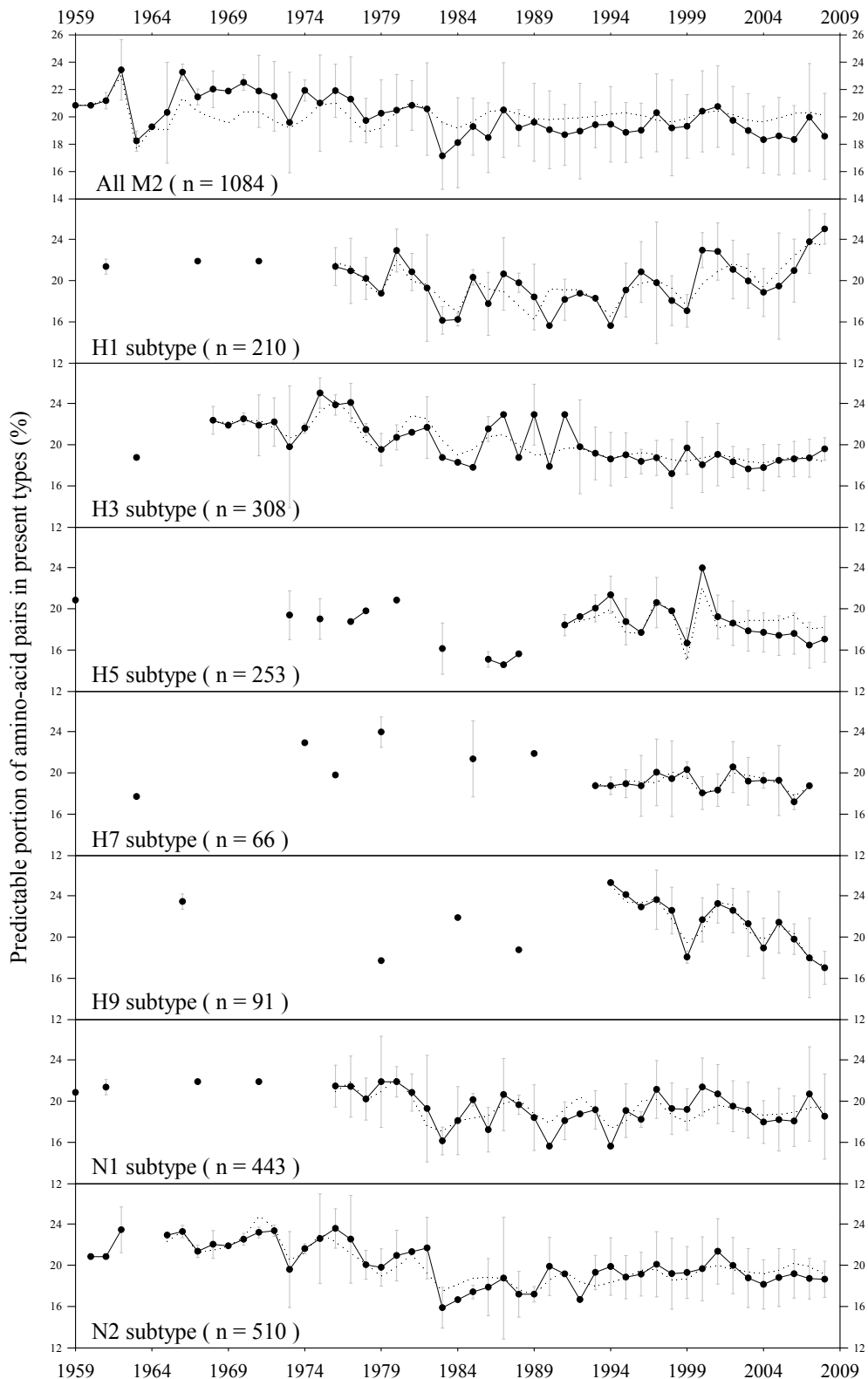These fluctuating solid curves suggested that the possi-

Figure 1. Evolutionary process of M2 proteins from influenza A viruses from 1959 to 2008 in terms of predictable portions with respect to different subtypes. The data are presented as mean±SD. The solid and dotted lines are the actual evolutionary process and fitted evolutionary process with analytical solution of differential equation.

bly analytical solution for *n* differential equations would be a sum of decaying exponential and sinusoidal functions $y(t) = \sum_{i=1}^{n} A_i e^{-k_i t} \cos(\alpha_i t + \varphi_i) + C$, where y was the fluctuating solid curve representing the predictable portion over time, A, α and k were parameters, t was time, φ was phase difference, and C was a constant [14].

3.2. Half-Life

This analytical solution governed a decaying trend with fluctuating solid curve because of negative exponential. Hence, we were able to determine the half-life of decaying phase of the fluctuating solid curve immediately. With decaying exponential, the half-life was $T_{1/2} = \frac{\ln(2)}{k} = \frac{0.696}{k}$, where $k = \frac{\ln(y_{peak}) - \ln(y_{trough})}{t_{interval}}$, which was the downhill half-life. Symmetrically, we were able

to compute the uphill half-life too because the uphill phase suggested that mutations led M2 proteins to become more predictable whereas the downhill phase suggested that mutations led M2 proteins to become less predictable.

Table 1 showed the computed half-life for all possible stratified peaks and troughs, and **Figure 2** compared the uphill half-life with the downhill one. As no statistical difference was found in **Figure 2**, it indicated that the uphill half-life was not different from the downhill half-life in principle.

These results suggested that we were able to use the analytical solution to fit the solid curve, because the unsolved problems in our previous study were that we were not able to determine the input function for this differential equation and were not able to determine if this evolutionary process was at steady-state.

Table 1. Half-life for all and each subtype of M2 proteins from influenza A viruses.

M2	Period	Length	Predictable portion (%)		Half-life (years)	
Subtype	Year	Years	Peak/Trough	Peak/Trough	Uphill	Downhill
All	1960-1962	3	20.83	23.44	18	
	1962-1963	2	23.44	18.23		6
	1963-1966	4	18.23	23.26	11	
	1966-1967	2	23.26	21.46		17
	1967-1970	4	21.46	22.51	58	
	1970-1973	4	22.51	19.59		20
	1973-1974	2	19.59	21.93	12	
	1974-1978	5	21.93	19.73		33
	1978-1981	4	19.73	20.83	51	
	1981-1983	3	20.83	17.14		11
	1983-1987	5	17.14	20.50	19	
	1987-1991	5	20.50	18.69		38
	1991-1997	7	18.69	20.30	59	
	1997-1998	2	20.30	19.19		25
	1998-2001	4	19.19	20.76	35	
	2001-2004	4	20.76	18.32		22
	2004-2007	4	18.32	19.97	32	
	2007-2008	2	19.97	18.58		19
H1	1976-1979	4	21.35	18.75		21
	1979-1980	2	18.75	22.92	7	
	1980-1983	4	22.92	16.15		8
	1983-1987	5	16.15	20.63	14	
	1987-1990	4	20.63	15.63		10
	1990-1992	3	15.63	18.75	11	
	1992-1994	3	18.75	15.63		11
	1994-1996	3	15.63	20.83	7	
	1996-1999	4	20.83	17.07		14
	1999-2000	2	17.07	22.94	5	
	2000-2004	5	22.94	18.85		18
	2004-2008	5	18.85	25.00	12	
H3	1972-1973	2	22.20	19.79		12
	1973-1975	3	19.79	25.00	9	
	1975-1979	5	25.00	19.53		14
	1979-1982	4	19.53	21.67	27	

|						
	1982-1985	4	21.67	17.80		14
	1985-1987	3	17.80	22.92	8	
	1987-1988	2	22.92	18.75		7
	1988-1989	2	18.75	22.92	7	
	1989-1990	2	22.92	17.90		6
	1990-1991	2	17.90	22.92	6	
	1991-1998	8	22.92	17.19		19
	1998-1999	2	17.19	19.68	10	
	1999-2003	5	19.68	17.65		32
	2003-2008	6	17.65	19.59	40	
H5	1994-1996	3	21.35	17.71		11
	1996-1997	2	17.71	20.60	9	
	1997-1999	3	20.60	16.67		10
	1999-2000	2	16.67	23.96	4	
	2000-2007	8	23.96	16.48		15
H7	1996-1999	4	18.75	20.31	35	
	1999-2000	2	20.31	18.06		12
	2000-2002	3	18.06	20.57	16	
	2002-2006	5	20.57	17.19		19
	2006-2007	2	17.19	18.75	16	
H9	1997-1999	3	23.62	18.06		8
	1999-2001	3	18.06	23.24	8	
	2001-2004	4	23.24	18.93		14
	2004-2005	2	18.93	21.42	11	
	2005-2008	4	21.42	17.01		12
N1	1976-1978	3	21.46	20.21		35
	1978-1979	2	20.21	21.88	18	
	1980-1983	4	21.88	16.15		9
	1983-1985	3	16.15	20.14	9	
	1985-1986	2	20.14	17.23		9
	1986-1987	2	17.23	20.63	8	
	1987-1990	4	20.63	15.63		10
	1990-1993	4	15.63	19.15	14	
	1993-1994	2	19.15	15.63		7
	1994-1997	4	15.63	21.13	9	
	1997-1999	3	21.13	19.18		22
	1999-2000	2	19.18	21.37	13	
	2000-2004	5	21.37	17.98		20
	2006-2007	2	18.07	20.70	10	
	2007-2008	2	20.70	18.52		13
N2	1961-1962	2	20.83	23.44	12	
	1966-1967	2	23.26	21.35		16
	1967-1972	6	21.35	23.34	47	
	1972-1973	2	23.34	19.59		8
	1973-1976	4	19.59	23.57	15	
	1976-1979	4	23.57	19.79		16
	1982-1983	2	21.67	15.89		4
	1983-1987	5	15.89	18.75	21	
	1987-1988	2	18.75	17.19		16
	1989-1990	2	17.19	19.89	10	
	1990-1992	3	19.89	16.67		12
	1992-1994	3	16.67	19.87	12	
	1994-1995	2	19.87	18.85		26
	1995-2001	7	18.85	21.36	39	
	2001-2004	4	21.36	18.15		17
	2004-2006	3	18.15	19.17	38	
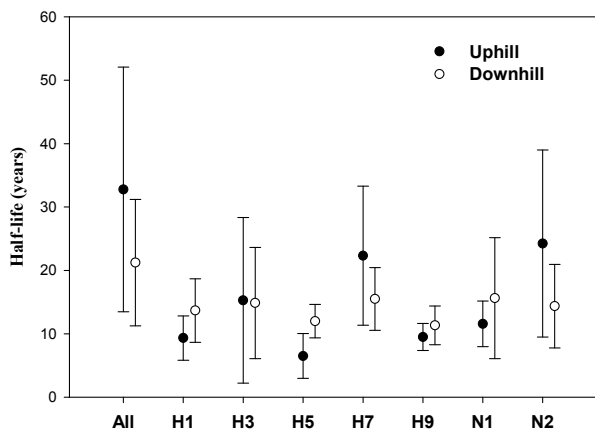
Figure 2. Comparison of uphill half-life with downhill half-life in all and different subtypes of M2 proteins from influenza A viruses. The data are presented as mean±SD.

3.3. Fitting

It was possible to use the analytical solution to fit the evolution of M2 proteins because the process of finding half-life provided the initial estimate for the parameter k_i in exponential terms.

The dotted lines in **Figure 1** were fitted lines using the analytical solution, and **Table 2** listed the fitted parameters for the analytical solution. As seen, the dotted lines generally were quite approximate to the evolutionary trend presented by the solid curve, indicating that the analytical solution was able to present the evolutionary process of M2 proteins from influenza A virus.

According to the general fitting principle, we were able to determine the goodness of our fitting through several ways, for example, 1) the Akaike's information criterion [15], 2) the plotting of residuals versus fitted predictable portion [16], 3) the plotting of residuals versus time [16], 4) the R or squared R value between fitted and actual data [13], etc. We mainly used the squared R value (**Table 2**) and Akaike's information criterion to determine if the difference between solid and dotted lines was acceptable. This was so because the sampled influenza A viruses were very unbalanced due to the practical difficulty in sampling, thus the solid lines could be biased on this account.

One possibility with this analytical solution was that the fluctuations would become less intensive as the time went on. This was possible because the evolutionary speed was becoming slower as less and less functional units needed to evolve. In fact, influenza viruses became more and more adapt to their environments after long-time evolution, thus they did not need to mutate a lot to suit for the changes in environments. This adaptation would lead the evolutionary speed of influenza A virus to be slower over time. For another example, the appendix in human could have very little speed for its evolution because its function is very much limited in general.

The use of differential equation to describe the evolution of proteins from influenza A viruses not only advanced our modeling ability in this field, but also provided us the tool to predict future mutations of influenza A viruses. For prevention of possible epidemic/pandemic, it is very important how to time mutations in proteins

Table 2. Parameters obtained after using the analytical solution to fit the evolutionary process of M2 proteins in **Figure 1.**

Subtype	All	H1	H3	H5	H7	H9	N1	N2
A_1	-1.0962	1.4086	-4.0515	-2.3573	1.5353	-3.9688	7.5911	-5.9992
k_1	0.0292	0.0000	0.1087	0.0762	0.0000	0.1360	0.2128	0.0677
a_1	0.9733	1.1999	1.0116	1.6356	1.2230	0.7309	0.6387	-0.1797
φ_1	-0.6025	-6.8726	1.1214	-1.9029	-2.5712	2.1098	3.8746	4.6835
A_2	60.1439	46.9288	-17.0142	-2.3072	-0.5694	-1.6362	2.1476	-2.0694
K_2	0.5698	0.0157	0.1206	0.0312	0.0000	0.0000	0.0557	0.0356
a_2	3.0209	6.3110	0.0685	-21.0433	1.9395	1.5054	1.5402	1.2495
φ_2	-10.7687	-4.0610	1.6425	-12.2725	3.0707	-3.1941	4.7661	0.7923
A_3	-2.5565	-1.4254	-2.6202	-1.2754	1.8712	3.3257	1.2562	1.1683
k_3	0.0687	0.0395	0.0589	0.0000	0.0000	0.0596	0.0198	0.0452
a_3	1.2499	2.4881	1.2671	2.8407	1.1729	0.2823	1.2257	4.2995
φ_3	-0.5466	3.2902	-2.6996	-0.1830	0.5379	4.8288	-1.5888	-10.0636
C	20.02	50.7341	18.4548	18.8510	19.1760	20.3824	18.9722	19.2776
R	0.5619	0.7951	0.8123	0.7793	0.8057	0.9601	0.7297	0.8772
R^2	0.3158	0.6322	0.6599	0.6074	0.6491	0.9217	0.5325	0.7695

from influenza A viruses. In the past we used the fast Fourier transform to do this job [3,5,17-20]. In the near future we are able to use the analytical solution with fitted parameters to time the mutations.

4. ACKNOWLEDGEMENTS

REFERENCES

[1] S. Yan and G. Wu. (2009) Describing evolution of hemagglutinins from influenza A viruses using a differential equation, Protein Pept. Lett., **16,** 794–804.

[2] G. Wu and S. Yan. (2002) Randomness in the primary structure of protein: Methods and implications, Mol. Biol. Today, **3,** 55–69.

[3] G. Wu and S. Yan. (2006) Mutation trend of hemagglutinin of influenza A virus: A review from computational mutation viewpoint, Acta Pharmacol. Sin., **27,** 513–526.

[4] G. Wu and S. Yan. (2006) Fate of influenza A virus proteins, Protein Pept. Lett., **13,** 377–384.

[5] G. Wu and S. Yan. (2008) Lecture notes on computational mutation, Nova Science Publishers, New York.

[6] T. Betakova. (2007) M2 protein-a proton channel of influenza A virus, Curr. Pharm. Des., **13,** 3231–3235.

[7] L. H. Pinto and R. A. Lamb. (2007) Controlling influenza virus replication by inhibiting its proton channel, Mol. Biosyst., **3,** 18–23

[8] J. Beigel and M. Bray. (2008) Current and future antiviral therapy of severe seasonal and avian influenza, Antiviral Res., **78,** 91–102.

[9] E. De Clercq and J. Neyts. (2007) Avian influenza A (H5N1) infection: Targets and strategies for chemotherapeutic intervention, Trends Pharmacol. Sci., **28,** 280–285.

[10] F. Hayden. (2009) Developing new antiviral agents for influenza treatment: What does the future hold? Clin. Infect. Dis., **48,** S3–S13.

[11] M. Schotsaert, M. De Filette, W. Fiers and X. Saelens. (2009) Universal M2 ectodomain-based influenza A vaccines: Preclinical and clinical developments, Expert Rev. Vaccines, **8,** 499–508.

[12] Influenza virus resources, (2009) http://www.ncbi.nlm.nih.gov/genomes/FLU/Database/multiple.cgi.

[13] SPSS Inc., (2002) SigmaPlot for Windows, Version 8.02.

[14] G. Wu. (1996) Fit fluctuating blood drug concentration: A beginner's first note, Pharmacol. Res., **33,** 379–383.

[15] K. Yamaoka, T. Nakagawa and T. Uno. (1978) Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacoketic equations, J. Pharmacokinet. Biopharm., **6**, 165–175.

[16] G. Wu, P. Cossettini and M. Furlanut. (1996) Prediction of blood cyclosporine concentrations in haematological patients with multidrug resistance by one-, two- and three-compartment models using Bayesian and nonlinear least squares methods, Pharmacol. Res., **34**, 47–57.

[17] G. Wu 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**, 143–148.

[18] G. Wu 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,** 70–78.

[19] G. Wu and 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.

[20] S. Yan and G. Wu. (2009) Prediction of mutation position, mutated amino acid and timing in hemagglutinins from North America H1 influenza A virus, J. Biomed. Sci. Eng, **2,** 190–199.