Fate of Neuraminidases of Influenza A Viruses

Shaomin Yan[1,a], Guang Wu[1*]

[1]National Engineering Research Center for Non-Food Biorefinery, State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Biomass Engineering Technology Research Center, Guangxi Key Laboratory of Bio-refinery
Guangxi Academy of Sciences
98 Daling Road, Nanning, Guangxi, 530007 China

Abstract. The current COVID-19 pandemic creates the biggest health and economic challenges to the world. However, not much knowledge is available about this coronavirus, SARS-CoV-2, because of its novelty. Indeed, it necessarily knows the fate of proteins generated by SARS-CoV-2. Anyway, before a large-scale study on proteins from SARS-CoV-2, it would be better to conduct a small-scale study on a well-known protein from influenza A viruses, because both are positive-sense RNA viruses. Thus, we applied a simple method of amino-acid pair probability to analyze 94 neuraminidases of influenza A viruses for better understanding of their fate. The results demonstrate three features of these neuraminidases: (i) the N1 neuraminidases are more susceptible to mutations, which is the current state of the neuraminidases; (ii) the N1 neuraminidases have undergone more mutations in the past, which is the history of the neuraminidases; and (iii) the N1 neuraminidases have a larger potential towards future mutations, which is the future of the neuraminidases. Moreover, our study reveals two clues on the mutation tendency, i.e. the mutations represent a degeneration process, and chickens, ducks and geese are rendered more susceptive to mutation. We hope to apply this approach to study the proteins from SARS-CoV-2 in near future.

1 Introduction

Influenza A viruses have led several pandemics in humans and animals [1]. The most famous pandemic is the Spanish flu, which led to a great loss of life and devastating impact on world economy before the ending of World War I [2, 3]. Then, there were a series of epidemics from 1957 until very recently around the world at different places with different species [4-7].

Heterogeneity is the characteristic of the original H5N1/01 isolates of genotypes A, C, D, and E [8–10]. Therefore, the huge genetic variability exists in the influenza A viruses, either by continuous and gradual mutation or by reassortment of gene segments between viruses, or both [3, 11]. A number of approaches have been used to study the antigenic drift and shift, for example, sequence analyses [12–14], the modeling of protein evolution [15], the mathematical model deals with both realistic epidemiological dynamics and viral evolution at the sequence level [16], and the traveling waves in a one-dimensional model [17]. Regarding the host-mediated variation of influenza A viruses, it has been so far proposed, at least, two mechanisms. The first mechanism focuses on the pressure of the antibody whereas the second mechanism emphases the selective pressure for the appearance of host cell variant with altered receptor binding specificities [18].

Because of this very big genetic variability in influenza A viruses, it imposes the difficulties to diagnosis, treatment, and vaccination development, so impedes prevention of influenza in humans to different degree. In reality, the genetic variability is not in the same range among influenza A virus proteins, because each type of proteins faces different selective pressure and plays different roles in virus. This leads some proteins to mutate faster than other proteins in a virus. Hence, it necessarily finds the pattern of mutation in each type of proteins for the benefit of diagnosis, treatment, vaccination and prevention of influenza in humans.

In influenza A virus, the mutations frequently occur in the RNA genes, which code for the surface glycoproteins, i.e. hemagglutinin and neuraminidase [18, 19]. The neuraminidase not only is involved in the binding of virus particles to receptors on host cells but also is the major antigen for neutralizing antibodies [20]. Of 15 subtype hemagglutinins of influenza viruses, the H5N1 viruses are highly pathogenic in view of classification of hemagglutinins. Furthermore, the A/teal/Hong Kong/W312/97 (H6N1) influenza virus and the human H5N1 and H9N2 influenza viruses contain similar genes encoding internal proteins [21, 22].

Since the enzymatic active centre in neuraminidase is the same across all influenza viruses, the blocking of this enzymatic active center is an approach to treat influenza. This could in principle stop the viral replication and alleviate, and even prevent the typical symptoms of influenza such as weakness, fever, and bodily aches and

[a]shaominyan@gxas.cn, hongguanglishibahao@gxas.cn

pains. As an enzyme, neuraminidase was discovered long time ago [23], cleaves the sialic acid residues terminally linked to glycoproteins and glycolipids, and thus becomes a major target for drugs and inhibitors [24].

In this study, we applied a simple method of amino-acid pair probability, which was developed by us [25], to estimate 94 neuraminidases from influenza A viruses for better understanding of their probabilistic fate.

2 Materials and Methods

The amino acid sequences of 94 neuraminidases are downloaded from the influenza virus resources [26].

2.1 Randomly Predictable Present Type of Amino-Acid Pair with Predictable Frequency

Neuraminidase Q807U2 has 54 serines (S) and 34 asparagines (N). The frequency of amino-acid pair of "SN" is 4 ($54/469 \times 34/468 \times 468 = 3.915$). Indeed, there are four "SN"s in neuraminidase Q807U2, so the type of "SN" is present and its frequency is 4. In this case, not only the presence of the type "SN" but also its frequency are predictable, and the difference between its actual and predicted values is 0.

2.2 Randomly Predictable Present Type of Amino-Acid Pair with Unpredictable Frequency

Neuraminidase Q807U2 has 46 glycines (G), and the frequency of "NG" is 3 ($34/469 \times 46/468 \times 468 = 3.335$). Interestingly, "NG" comes out 9 times in neuraminidase Q807U2. Naturally, the presence of the type "NG" is predictable, but its frequency is unpredictable, and the difference between its actual and predicted values is 6.

2.3 Randomly Unpredictable Present Type of Amino-Acid Pair

Neuraminidase Q807U2 has 14 glutamines (Q) and 10 histidines (H). The frequency of "QH" is 0 ($14/469 \times 10/468 \times 468 = 0.299$), so the type of "QH" would not come out in neuraminidase Q807U2. By contrast, the "QH" appears twice in the reality, so the presence of the type of "QH" is unpredictable. As a result, its frequency is unpredictable too, and the difference between its actual and predicted values is 2.

2.4 Randomly Predictable Absent Type of Amino-Acid Pair

Neuraminidase Q807U2 has 7 methionines (M), and the frequency of "MQ" is 0 ($7/469 \times 14/468 \times 468 = 0.209$), consequently the type of "MQ" would not show in neuraminidase Q807U2, which is true. Subsequently, the absence of the type of "MQ" with its frequency is predictable, and the difference between its actual and predicted values is 0.

2.5 Randomly Unpredictable Absent Type of Amino-Acid Pairs

The frequency of "GG" is 4 ($46/469 \times 45/468 \times 468 = 4.414$), i.e. four "GG"s exist in neuraminidase Q807U2. But no "GG" is found. As such, the absence of "GG" is unpredictable. Fairly its frequency is unpredictable too, and the difference between the actual and predicted values is –4.

2.6 Statistics

For actual and predicted values in a single protein, the statistical inference is performed as follows. Essentially each of 20 kinds of amino acids has the chance of 1/20 ($p = 0.05$) to repeat again, and a type of amino-acid pair has the chance of 1/400 ($p = 0.0025$) to repeat again. In neuraminidase Q807U2, the most abundant amino acid is 54 "S"s, and the least abundant amino acid is 7 "M"s, If the first amino acid is "S", then the chance of the second amino acid being "S" is 53/468 ($p = 0.113 > 0.05$). If the first amino acid is "M", then the chance of the second amino acid being "M" is 6/468 ($p = 0.013 < 0.05$). Following that, the chance of the first amino-acid pair being "SS" is $54/469 \times 53/468$ ($p = 0.013 < 0.05$), and the chance of the second amino-acid pair being "SS" is $52/467 \times 51/466$ ($p = 0.012 < 0.05$). For the least abundant amino acids "M", the chance of the first amino-acid pair being "MM" is $7/469 \times 6/468$ ($p = 0.0002 < 0.001$), and the chance of the second amino-acid pair being "MM" is $5/467 \times 4/466$ ($p = 0.0001 < 0.001$). Reasonably, the probability is less than 0.05 if the difference between actual and predicted values is greater than or equal to one.

For comparisons among proteins, the statistical inference is done as follows. The Kolmogorov-Smirnov test is used to examine all the data to determine their distribution property. For the normal distribution, the data are shown as mean±SD. For the non-normal distribution, the data are shown as median with interquartile range. The outlier is detected based on the Healy's method [27]. The one-way ANOVA and the Friedman ANOVA rank tests are used for parametric and non-parametric tests, respectively, followed by comparison tests. The SigmaStat for Windows (SPSS Inc, 1992–2003) is used to operate all the statistical tests, and the $p < 0.05$ is considered statistically significant.

3 Results

After all the computation, the amino-acid pairs in a neuraminidase are classified into predictable and unpredictable portions. After comparison of the percentages of predictable portion with unpredictable portion among different neuraminidases, one can know which neuraminidase has a larger unpredictable portion than other neuraminidases. It turns out this neuraminidase would be more susceptible to mutations accordingly [28–34].

Fig. 1 illustrates the predictable and unpredictable portions in three subtypes of neuraminidases. In Fig. 1, the length of each bar represents 100%, which is located at

both unpredictable and predictable sites separated by dotted line. For example, the absent types of N1 neuraminidases (the dark green bar) are composed of 31.75% predictable portion (the right panel) and 68.25% unpredictable portion (the left panel). Statistical differences are found in comparison between two subtype neuraminidases in terms of the absent types ($p < 0.001$), while no significant differences are in both the present types and frequencies.
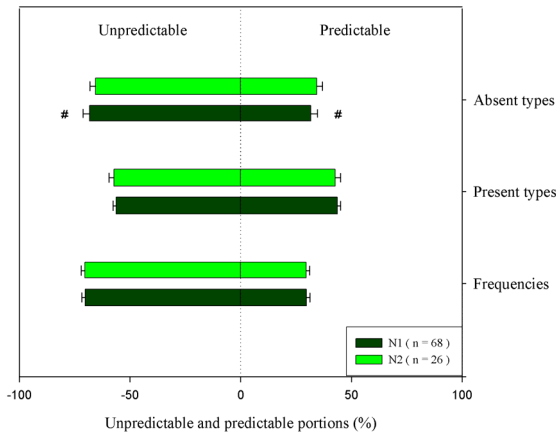

Fig1. Predictable and unpredictable portions of amino-acid pairs in neuraminidases. The data are shown as mean±SD. #, the statistical difference between two groups at $p < 0.001$ level

Hereafter, attention is paid to the unpredictable portions (the left panel in Fig. 1), because they are not driven by randomness. As discussed in Materials and Methods section, an unpredictable portion includes both unpredictable types and predictable types with unpredictable frequencies. These data can be in forms that the actual values are either larger or smaller than the predicted values. In the past, we demonstrated that the amino-acid pairs whose actual value is larger than their predicted value are likely to be targeted by mutations and the amino-acid pairs whose actual value is smaller than their predicted one are more likely to be formed through mutations [28–34].
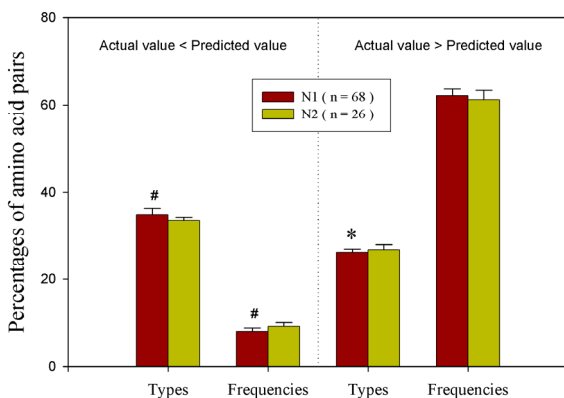

Fig2. Percentages of unpredictable types and frequencies in terms of whether the actual value is larger or smaller than the predicted value in neuraminidases. The data are shown as mean±SD. * and #, the statistical difference between two groups at $p < 0.05$ and $p < 0.001$ level, respectively

Fig. 2 depicts the percentages of unpredictable types and frequencies in terms of whether the actual values are larger or smaller than their predicted values in two subtypes of neuraminidases. Practically Fig. 2 is a subdivision of Fig. 1 with stratification of the data in the left panel of Fig. 1 into two criteria, i.e. the actual values are smaller than their predicted values or vice versa. The left panel of Fig. 2 details that the unpredictable types are statistically larger in N1 neuraminidases than in the N2 neuraminidases, and the unpredictable frequencies are statistically smaller in N1 neuraminidases than in the N2 neuraminidases. Meanwhile the right panel of Fig. 2 details that the unpredictable types are statistically smaller in N1 neuraminidases than in N2 neuraminidases. In the wake of the data in the right panel of Fig. 2, we understand that the N1 neuraminidases have undergone more mutations in the past. In the wake of the data in the left panel of Fig. 2, it turns out that the mutation pattern in N1 neuraminidases is to form the new type of amino acid pairs rather than to increase the frequency of present type of amino acid pairs.

Along these results, attention is particularly directed to the difference between the actual and predicted values for we have shown that the larger the difference between the actual and predicted values is, the bigger the potential towards future mutations is [28–34].
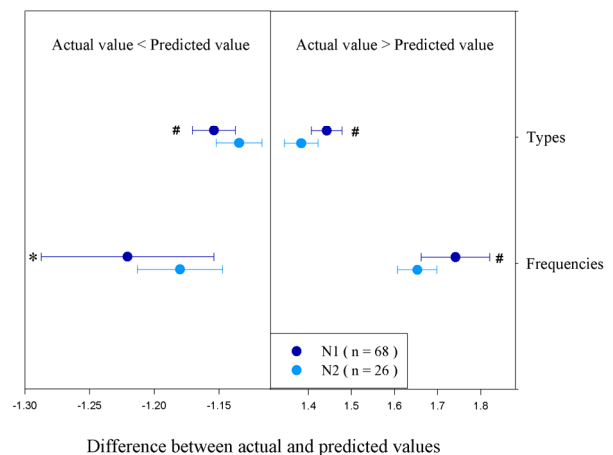

Fig3. Difference between the actual and predicted values in N1 and N2 neuraminidases. The data are shown as mean±SD. * and #, the statistical difference between N1 and N2 neuraminidases at $p < 0.05$ and $p < 0.001$ level, respectively

Fig. 3 displays the size of the difference between the actual and predicted values in two subtypes of neuraminidases. Apparently, the N1 neuraminidases have larger differences than the N2 neuraminidases, suggesting that the N1 neuraminidases are more vulnerable to future mutations than the N2 neuraminidases.
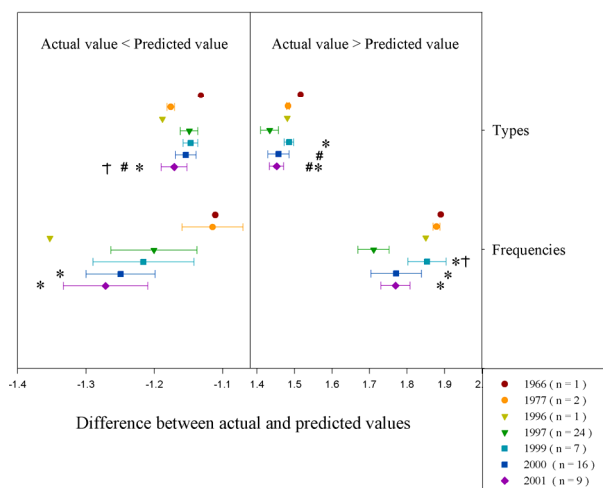
Fig4. Difference between the actual and predicted values in N1 neuraminidases isolated from different years. The data are shown as mean±SD. *, # and †, the statistical difference from 1997, 1999 and 2000 groups, respectively, at $p < 0.05$ level

Yet, the size of difference between the actual and predicted values can implicate the direction of future mutations if we arrange the size of difference along the time course. Fig. 4 pictures the size of difference between the actual and predicted values of N1 neuraminidases of influenza A viruses isolated along the time course. General speaking, the differences increase in the amino-acid pairs whose actual value is smaller than their predicted value (the left panel in Fig. 4), whereas the differences decrease in the amino-acid pairs whose actual value is larger than their predicted value (the right panel in Fig. 4). As the percentage of frequency is about 6-fold higher in the right panel than in the left panel (Fig. 2), these results imply that a mutation in the N1 neuraminidases is to reduce the difference between the actual and predicted values, so that the construction of amino-acid pairs becomes more random.

Additionally, the difference between actual and predicted values can indicate which species is subject more to mutations if we arrange the number of amino-acid pairs against the difference between actual and predicted values in neuraminidases from different species.

Fig. 5 exhibits this type of analysis, where the scale of the vertical axes is presented by logarithm in order to stress the amino-acid pairs with large differences between actual and predicted values. As can be seen in Fig. 5, the N1 neuraminidases of influenza A viruses from chickens, ducks and geese have the largest difference between actual and predicted values.
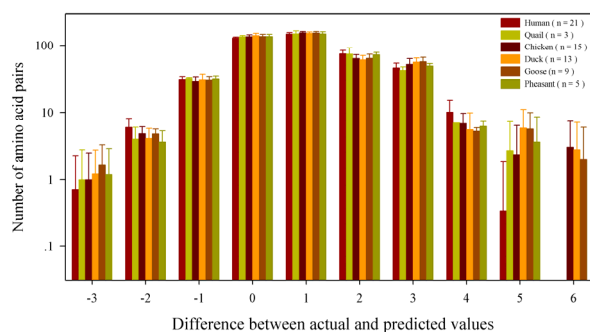

Fig5. Number of amino-acid pairs of neuraminidases from different species with respect to the difference between their actual and predicted values. The data are shown as mean±SD

4 Discussion

This study demonstrates three features from N1 and N2 neuraminidases of influenza A viruses. Besides, this study generates two clues on the mutation tendency and on the species susceptibility.

The first feature is that the N1 neuraminidases are more susceptible to mutations, which is the current state of the neuraminidases. This feature is supported by the fact that the unpredictable portion of absent types is larger in the N1 neuraminidases than in the N2 neuraminidases, although the percentages of the present types and frequencies are similar in the two subgroups of neuraminidases (Fig. 1).

The second feature is that the N1 neuraminidases have undergone more mutations in the past, which is the history of the neuraminidases. The reasoning is that the N1 neuraminidases have a larger percentage of unpredictable types with a smaller percentage of unpredictable frequencies (the right panel in Fig. 2).

The third feature is that the N1 neuraminidases have a stronger potential towards future mutations, which is the future of the neuraminidases. The interpretation is that the N1 neuraminidases have larger differences between actual and predicted values (Fig. 3). This implies that the N1 neuraminidases have a stronger potential towards future mutations.

From a probabilistic viewpoint, the mutation tendency hints that at least some mutations direct to a probabilistically capable way. In order to make a protein functional, nature should deliberately spend more time and energy to construct particular amino-acid pairs with big differences between their actual and predicted frequencies while the random construction of an amino-acid pair is the least time- and energy-consuming [23]. The future mutations in neuraminidases are likely to link with the balance between the random and purpose-oriented construction of amino-acid pairs. Therefore, some mutations on neuraminidases represent a degeneration process in Fig. 4, and this observation is identical with the proteins, which have a high year-to-year mutation rate [28–34].

From a point of view of species susceptibility, the second hint is that chickens, ducks and geese are rendered more susceptive to mutation. The species susceptibility

probabilistically relies on the number of amino-acid pairs with the biggest difference between actual and predicted values. Fig. 5, at least partially, highlights why so many mutations were found in the neuraminidases of influenza A viruses from chickens, ducks and geese. Fig. 5 supports the finding that avian species plays an important role in harboring a large amount of influenza A virus strains, which can contribute genes to form potentially new pandemic human strains [1]. Also, avian viruses can infect humans without acquiring human influenza genes by reassortment in an intermediate host [35].

In conclusion, the fate of neuraminidases from influenza A viruses could shed some light on the current COVID-19 pandemic.

ACKNOWLEDGMENTS

REFERENCES

1. R.D. de Vries, S. Herfst, M Richard (2018). Avian influenza A virus pandemic preparedness and vaccine development. Vaccines (Basel) 6, 46.

2. E. Tognotti (2003). Scientific triumphalism and learning from facts: bacteriology and the "Spanish flu" challenge of 1918. Social History of Medicine 16, 97–110.

3. A.H. Reid, J.K. Taubenberger (2003). The origin of the 1918 pandemic influenza virus: a continuing enigma. Journal of General Virology 84, 2285–2292.

4. Y.P. Lin, M. Shaw, V. Gregory, K. Cameron, W. Lim, A. Klimov, K. Subbarao, Y. Guan, S. Krauss, K. Shortridge, R. Webster, N. Cox, A. Hay (2000). Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. Proceedings of the National Academy of Sciences of the United States of America 97, 9654–9658.

5. M. Hatta, P. Gao, P. Halfmann, Y. Kawaoka (2001). Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science 293: 1840–1842.

6. T. Noda, Y. Kawaoka (2010). Genetic and molecular characteristics of the (2009 H1N1 pandemic influenza A virus. Nippon Rinsho. Japanese Journal of Clinical Medicine 68, 1590-1593.

7. G. Tsoucalas, M. Sgantzos (2016). The 2009 influenza A virus subtype H1N1 pandemic, a glance from Greece. Le Infezioni in Medicina 24, 259-264.

8. G.W. Wood, J Banks, J.W. McCauley, D.J. Alexander (1994). Deduced amino acid sequences of the haemagglutinin of H5N1 avian influenza virus isolates from an outbreak in turkeys in Norfolk, England. Archives of Virology 134, 185–194.

9. A.S. Lipatov, S. Krauss, Y. Guan, M. Peiris, J.E. Rehg, D.R. Perez, R.G. Webster (2003). Neurovirulence in mice of H5N1 influenza virus genotypes isolated from Hong Kong poultry in 2001. Journal of Virology 77, 3816–3823.

10. Y. Guan, J.S.M. Peiris, A.S. Lipatov, T.M. Ellis, K.C. Dyrting, S. Krauss, L.J. Zhang, R.G. Webster, K.F (2002). Shortridge. Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. Proceedings of the National Academy of Sciences of the United States of America 99, 8950–8955.

11. Y. Guan, M. Peiris, K.F. Kong, K.C. Dyrting, T.M. Ellis, T. Sit, L.J. Zhang, K.F. Shortridge (2002). H5N1 influenza viruses isolated from geese in Southeastern China: evidence for genetic reassortment and interspecies transmission to ducks. Virology 292, 16–23.

12. C. Bender, H. Hall, J. Huang, A. Klimov, N. Cox, A. Hay, V. Gregory, K. Cameron, W. Lim, K. Subbarao (1999). Characterization of the surface proteins of influenza A (H5N1) viruses isolated from humans in 1997-1998. Virology 254, 115–123.

13. L.V. Gubareva, D.V. Novikov, F.G. Hayden (2002). Assessment of hemagglutinin sequence heterogeneity during influenza virus transmission in families. Journal of Infectious Diseases 186, 1575–1581.

14. R.G. Webster, Y. Guan, M. Peiris, D. Walker, S. Krauss, N.N. Zhou, E.A. Govorkova, T.M. Ellis, K.C. Dyrting, T. Sit, D.R. Perez, K.F. Shortridge (2002). Characterization of H5N1 influenza viruses that continue to circulate in Geese in Southeastern China. Journal of Virology 76, 118–126.

15. M.W. Deem, H.Y. Lee (2003). Sequence space localization in the immune system response to vaccination and disease. Physical Review Letters 91, 068101–068104.

16. N.M. Ferguson, A.P (2003). Galvani, Bush RM. Ecological and immunological determinants of influenza evolution. Nature 422, 428–433.

17. J, Lin, V. Andreasen, R. Casagrandi, S.A. Levin (2003). Traveling waves in a model of influenza A drift. Journal of Theoretical Biology 222, 437–445.

18. Y. Suzuki (2001). Host mediated variation and receptor binding specificity of influenza viruses. Advances in Experimental Medicine and Biology 491, 445–451.

19. M.R. Hilleman (2002). Realities and enigmas of human viral influenza: pathogenesis, epidemiology and control. Vaccine 20, 3068–3087.

20. M.C. Zambon (1999). Epidemiology and pathogenesis of influenza. Journal of Antimicrobial Chemotherapy 44 Suppl. B, 3–9.

21. E. Hoffmann, J. Stech, I. Leneva, S. Krauss, C. Scholtissek, P.S. Chin, M. Peiris, K.F. Shortridge, 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, 6309–6315.

22. P.S. Chin, E. Hoffmann, R. Webby, R.G. Webster, Y. Guan, M. Peiris, K.F. Shortridge (2002). Molecular

evolution of H6 influenza viruses from poultry in Southeastern China: prevalence of H6N1 influenza viruses possessing seven A/Hong Kong/156/97 (H5N1)-like genes in poultry, Journal of Virology 76, 507–516.

23. G.K. Hirst (1942). Absorption of influenza haemagglutinins and virus by red blood cells. Journal of Experimental Medicine 76, 195–209.

24. C.U. Kim, X. Chen, D.B. Mendel (1999). Neuraminidase inhibitors as anti-influenza virus agents. Antiviral Chemistry and Chemotherapy 10, 141–54.

25. G. Wu, S. Yan 2008 *Lecture Notes on Computational Mutation*. Nova Science Publishers, New York.

26. The National Center for Biotechnology Information. Influenza Virus Resources (2020). https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database.

27. M.J.R. Healy (1979). Outliers in clinical chemistry quality control schemes. Clinical Chemistry 25 675–677.

28. S. Yan, G. Wu (2010). Mutation pattern in human adrenoleukodystrophy protein in terms of amino-acid pair predictability. J Biomed Sci Eng 3, 262–7

29. S. Yan, G. Wu (2010). Creation and application of computational mutation. Journal of Guangxi Academy of Sciences 26, 130–139.

30. S. Yan, G. Wu (2013). Small variations between species/subtypes attributed to reassortment evidenced from polymerase basic protein 1 with other seven proteins from influenza A virus. Transboundary and Emerging Diseases 60, 110–119.

31. Q. Yu, S. Yan, G (2013). Wu Mutation patterns of human proto-oncogene tyrosine-protein kinase receptor RET by means of amino-acid pair predictability. Life Science Journal 10, 1667–1672.

32. S. Yan, G. Wu (2013). Possibility of cross-species/subtype reassortments in influenza A viruses: An analysis on nonstructural protein variations. Virulence 4, 716–725.

33. S.Y. Long，S. Yan, G. Wu (2014). Mutation patterns in human cystic fibrosis transmembrane conductance regulator protein (in Chinese). Guangxi Sciences 21, 671–676.

34. C.H. Zhang, W.S. Ye, S. Yan, G. Wu (2019). Mutation patterns in lysostaphin. Journal of Biomedical Science and Engineering 12, 322–332.

35. K. Subbarao, J. Katz (2000). Avian influenza viruses infecting humans. Cellular and Molecular Life Sciences 57, 1770–1784.