King Saud University

Saudi Journal of Biological Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Detecting early-warning signals for influenza A pandemic based on protein dynamical network biomarkers

CrossMark

Jie Gao [a,b,]*, Kang Wang [a], Tao Ding [a]

[a] *School of Science, Jiangnan University, Wuxi 214122, China*
[b] *Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China*

Abstract The outbreak of influenza A comes from a relatively stable state is a critical phenomenon on epidemic. In this paper, influenza A varying from different states is studied in the method of dynamical network biomarkers (DNB). Through studying DNB of influenza A virus protein, we can detect the warning signals of outbreak for influenza A and obtain a composite index. The composite index varies along with the state of pandemic influenza, which gives a clue showing the turn point of outbreak. The low value (<1) steady state of the composite index means influenza A is normally in the relatively steady stage. Meanwhile, if the composite index of a certain year increases by more than 0.8 relative to the previous year and it is less than 1 and it increases sharply and reaches a peak being larger than 1 in next year, it means the year is normal in the critical state before outbreak and the next year is normally in the outbreak state. Therefore, we can predict the outbreak of influenza A and identify the critical state before influenza A outbreak or outbreak state by observing the variation of index value.

© 2017 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

It is proved that there exists a kind of common critical phenomenon in many complex biological processes, i.e. a relative stable state enters into another state quickly after a soon critical point (Chen et al., 2012; Liu et al., 2012) There is the kind of critical phenomenon for influenza A, because it needs only a very short period of time quickly from a relative stable state to outbreak state after a critical point. Thus in order to prevent and control the outbreak of influenza A pandemic timely and effectively, the key solution lies in predicting the critical point before the outbreak.

At present, influenza A is studied from many aspects. Ya-Nan Pan et al. found that the spatio-temporal network that connects the cities with human cases along the order of outbreak timing emerges two-section-power-law edge-length distribution, using the empirical analysis and modeling studies

---

ELSEVIER | **Production and hosting by Elsevier**

(Pan et al., 2014; Zhang, 2016). Chang et al. (2009) studied the vaccine for influenza, so as to achieve the effect of prevention of influenza. Banerjee et al. (2015) made full comparisons for the structural features of all H1N1 HA gene sequences and the composition of global amino acid to make it possible to depict the developing trend of influenza A. He et al. (2014) also made indepth studies to identify HA protein epitopes of avian influenza virus.

This paper studies the different states of influenza A using dynamical network biomarkers (DNB). Through studying DNB of influenza A virus protein, we can detect the warning signals of outbreak for influenza A and obtain a composite index. The composite index varies along with the state of pandemic influenza, which gives a clue showing the turn point of outbreak. Therefore, we can predict the outbreak of influenza A and identify the critical state before influenza A outbreak or outbreak state by observing the variation of index value. This indicates the composite index can provide reliable and significant warning information to detect the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic.

2. Method

The concept of network biomarkers is set up with the development of high-throughput genomic technologies and the systematic and multidimensional study of molecular expression profiling (Liu et al., 2014; Wu et al., 2012). This concept refers to a series of markers as well as their mutual relations and has been proposed as a new marker type (Jin et al., 2008; Yao et al., 2015). Compared with traditional biomarkers, these markers can accurately distinguish disease states for taking the links between the molecules into consideration (Simon, 2005; Ludwig and Weinstein, 2005). However, it is used to diagnose the states of diseases, not for the detecting the critical point before the outbreak of diseases.

The method of dynamic network biomarkers focuses on the detection and assessment of different stages of the disease in the development of disease. This is a time-dependent method (Sun et al., 2014). It studies the location changes of the markers over time and the relationship among network markers over time changing. Meanwhile, this method can construct three-dimensional images showing the interaction relationship between the markers. Therefore the study of Network markers focuses on the molecular interactions and distinguishes normal and disease states. The study of dynamic network markers focusing on dynamic changes, is helpful to discover the marker accurately, comprehensively, and further to distinguish the state of disease before outbreak. It does not only depend on the method of small sample excavation mode markers, but also make it easier for clinical application. At the same time it can be used in wide studies to find early warning signals in any biological process, such as differentiation, senescence and cell cycle of each phase as well as key change.

3. Results

3.1. Data

Here are ten of proteins for influenza A virus hemagglutinin (HA), matrix protein, matrix protein 2, neuraminidase, non-structural protein 1, non-structural protein 2, nucleocapsid protein, PA RNA polymerase, PB1 RNA polymerase and PB2 RNA polymerase. They are composed of 20 different amino acids link to form polymers. This paper selects influenza A virus protein sequences from 1933 to 2015 from the NCBI website (www.ncbi.nlm.nih.gov/), whereas some data in 1937, 1938, 1939, 1940, 1941, 1942, 1944, 1951, 1952, 1953, 1954 and 1955 years are absent.

3.2. Model

3.2.1. Defining dynamic network biomarker

Taking HA protein as an example firstly, we suppose that a HA protein marked y is linked sequentially by t numbers of amino acids. Its amino acid sequence is represented by $y = x_1x_2\cdots x_t$, in which $x_i \in$ {A, V, L, I, P, F, W, M, D, E, G, S, T, C, Y, N, Q, K, R, H}; $i = 1,2,\ldots,t$. We suppose s-1-th year have m numbers of influenza virus HA proteins all over the world and its amino acid sequence is represented by $y_{s-1,1}$, $y_{s-1,2},\ldots,y_{s-1,m}$. Meanwhile, We suppose s-th year have n numbers of influenza virus HA proteins all over the world and its amino acid sequence is represented by $y_{s,1}$, $y_{s,2},\ldots,y_{s,n}$. The amino acid number of the $y_{i,j}$ is marked $c_{i,j}$, where $i = $ s−1,s; $j = 1,2,\ldots,q$; $q = \max\{m,n\}$. Sequentially selecting the i-th amino acid for $y_{s-1,1},y_{s-1,2},\ldots,y_{s-1,m}$ to form a new amino acid sequence is defined $Z_{s-1,i}$, and then take out the largest one of amino acids number. If the maximum number of amino acids has two or more than two, we take the first amino acid without loss of generality. At the same time, these amino acids are marked x_i, where $i = 1,2,\ldots,k$; $k = \max\{c_{s-1,1},c_{s-1,2},\ldots,c_{s-1,m}\}$. We individually connect them in order to form a new amino acid sequence ($U_{S-1} = x_1x_2\ldots x_k$) and then separately compare with corresponding amino acids of $y_{s,1},y_{s,2},\ldots,y_{s,n}$ one by one. If they are different, the assignment is one, on the contrary the assignment is zero. Therefore, n new sequences are represented by $E_{s,1},E_{s,2},\ldots,E_{s,n}$ are obtained in s-th year. Then we calculate their mean (M), standard deviation (SD) and coefficient of variation (CV). Their computation formulas are as follows:

$$M_s = \frac{\sum_{i=1}^{n} f(s,i)}{n} \tag{1}$$

$$SD_s = \sqrt{\frac{\sum_{i=1}^{n}(f(s,i) - M_s)^2}{n}} \tag{2}$$

$$CV_s = \frac{SD_S}{M_s} \tag{3}$$

where $f(s,i)$ represents the frequency of occurrence of one in sequence $E_{s,i}$. Similarly, we calculate M, SD and CV of the other nine proteins. The protein that values of CV_s are the top three are defined as core protein (CP), and the others are no-core protein (NP). CP is a set of high confidence interactions of proteins, which forms a sub-network called influenza A virus proteins of the protein dynamical network biomarkers.

3.2.2. The early warning model for influenza A

We calculate the frequencies of the 20 kinds of amino acids, and the computation formulas are as follows:

$$f_{x_i}(s) = \frac{\sum_{j=1}^{n} f_{x_i}(s,j)}{n} \tag{4}$$

where $f_{x_i}(s,j)$ represents the frequency of occurrence of amino acid x_i in amino acid sequence $y_{s,j}$. Now, we can get a 23 dimensional characteristic value vector of HA protein. By the same way, the $f_{x_i}(s)$ of the other nine proteins can be calculated in turn, so we can get a characteristic value matrix ($X = [V_1(s), V_2(s), \ldots, V_{10}(s)]$), where $V_t(s)$ represents the characteristic value vector of the t-th influenza A protein, $t = 1, 2, \ldots, 10$. Defining the characteristic distance between proteins:

$$d_w = \sqrt{(M_s - M_{ws})^2 + (\sigma_s - \sigma_{ws})^2 + (CV_s - CV_{ws})^2 + \sum_{i=1}^{20}(f_{x_i}(s) - f_{wx_i}(s))^2} \quad (5)$$

where v and w represents the v-th and the w-th protein respectively.

The core proteins are not only the universal indicators to detect the complex outbreak signal of influenza A, but also the dominant or driving network of the whole protein system in the development, mutation and outbreak of the critical stages. In fact, the dominant network breaks through the limits of variation in the first time, first enters to the state of variation, and then affects other proteins and lead to the transfer of the entire system. Therefore, the determination of the dominant network can not only detect system in the critical state before break out, also help to reveal the underlying mechanism of influenza A virus proteins from the dimension of dynamic network. By combining the above properties of the core proteins, we can define a composite index:

$$I = \frac{\overline{CV_k} \cdot \overline{CP_{cd}}}{\overline{NP_{cd}}} \quad (6)$$

where $\overline{CV_k}$ represents the average value of the core proteins' CV_s, $\overline{CP_{cd}}$ is the average value of the characteristic distance between the core proteins, and $\overline{NP_{cd}}$ is the average value of the characteristic distance between the core and non-core proteins.

When $I_s - I_{s-1} \geqslant 0.08$, and $I_{s-1} < 1$, $I_s < 1$, $I_{s+1} \geqslant 1$, it can be concluded that s-th year is in the critical state before the outbreak, the $s + 1$ year is in the outbreak state.

Although the amino acid sequence of each protein will fluctuate randomly, the composite index can provide a credible and significant early warning when the influenza A virus is close to the critical state before the outbreak or the outbreak state.

3.2.3. Data processing

As shown in Table 1, we can calculate the composite index of the 1934–2015 using the above methods. However, we can't figure out the composite index of some years, because some data in 1937, 1938, 1939, 1940, 1941, 1942, 1944, 1951, 1952, 1953, 1954 and 1955 years are absent.

4. Discussion

The dynamic network markers of Pandemic influenza virus vary in the whole process from a relatively stable state to the critical state before outbreak as well as the outbreak state, which results in the status transfer of the entire network and finally results in fluctuations in the composite index. Therefore, by observing the transformation of the composite index, we

Table 1 Composite index values from 1934 to 2015.

Year	I	Year	I	Year	I
1934	0.723227	1970	0.627428	1993	1.187957
1935	0.962083	1971	0.728516	1994	1.225976
1936	0.43403	1972	2.379322	1995	0.164168
1943	0.543059	1973	0.79888	1996	0.611397
1946	0.441866	1974	0.527835	1997	0.711091
1947	0.851604	1975	0.801294	1998	0.629408
1948	0.448092	1976	2.275519	1999	0.781102
1949	0.760293	1977	2.182157	2000	0.710281
1950	0.734349	1978	0.438169	2001	0.295353
1951	1.027341	1979	0.32697	2002	0.660193
1957	0.859728	1980	0.746082	2003	0.405805
1958	0.949281	1981	0.455632	2004	0.45421
1959	0.474866	1982	0.650454	2005	0.772595
1960	0.550811	1983	0.449789	2006	1.595902
1961	0.708772	1984	0.354632	2007	0.476057
1962	0.08078	1985	0.939702	2008	0.798138
1963	0.980012	1986	1.166947	2009	1.344778
1964	0.650854	1987	0.655971	2010	0.962241
1965	0.527201	1988	1.033681	2011	0.440067
1966	0.580452	1989	0.912527	2012	0.635735
1967	0.666783	1990	1.898656	2013	0.606009
1968	2.31271	1991	1.248818	2014	0.806321
1969	1.081257	1992	1.032401	2015	2.516147

can predict the critical state before the outbreak of pandemic influenza and the outbreak state.

The flu broke out in Hong Kong in 1968 and continued until 1969, 7.5 million people died during that period. In 1972, influenza broke out in Henan Province and quickly spread to the entire province. As shown in Fig. 1, in 1965, the composite index value is 0.527201; 1966 is 0.580452; 1967 is 0.666783; 1968 is 2.31271; 1969 is 1.081257; 1970 is 0.405805; 1971 is 0.728516 and 1972 is 2.379322. Because $I_{1967} - I_{1966} = 0.086331 > 0.08$, $I_{1966} < 1$, $I_{1967} < 1$, $I_{1968} \geqslant 1$, $I_{1969} \geqslant 1$, the state in 1967 is the critical state before the outbreak, the state in 1968 is the outbreak state and the state continues until 1969. Similarly, $I_{1971} - I_{1972} =$
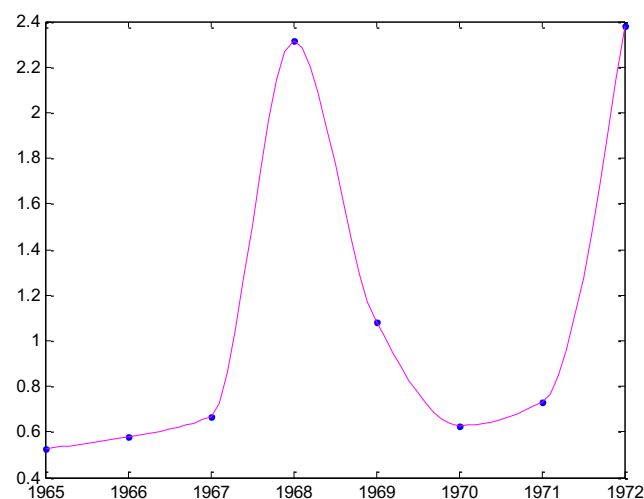

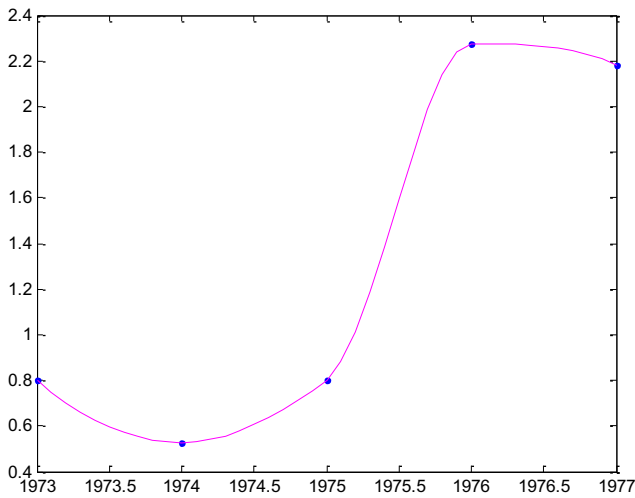**Figure 1** Trend Chart of composite index values from 1965 to 1972.

Figure 2 Trend Chart of composite index values from 1973 to 1977.

Figure 4 Trend Chart of composite index values from 2003 to 2009.

0.101088 > 0.08, $I_{1970} < 1$, $I_{1971} < 1$, $I_{1972} \geqslant 1$, so the state in 1971 is the critical state before the outbreak, the state in 1972 is the outbreak state.

The influenza A broke out in The United States, Russia and Japan in 1976 and 1977. The incidence rate was very high in young people. As shown in Fig. 2, in 1973, the composite index value is 0.79888; 1974 is 0.527835; 1975 is 0.801294; 1976 is 2.275519; 1977 is 2.182157. $I_{1975} - I_{1974} = 0.273459 > 0.08$, $I_{1974} < 1$, $I_{1975} < 1$, $I_{1976} \geqslant 1$, $I_{1977} \geqslant 1$, so the state in 1975 is the critical state before the outbreak, the state in 1976 is the outbreak state.

The influenza A broke out in The United States and Japan in 1986. Meanwhile, many countries in Asia and Europe had the outbreak of influenza A. As shown in Fig. 3, in 1983, the composite index value is 0.449789; 1984 is 0.354632; 1985 is 0.939702; 1986 is 1.166947. $I_{1985} - I_{1984} = 0.58507 > 0.08$, $I_{1984} < 1$, $I_{1985} < 1$, $I_{1986} \geqslant 1$, so the state in 1985 is the critical state before the outbreak, the state in 1986 is the outbreak state.
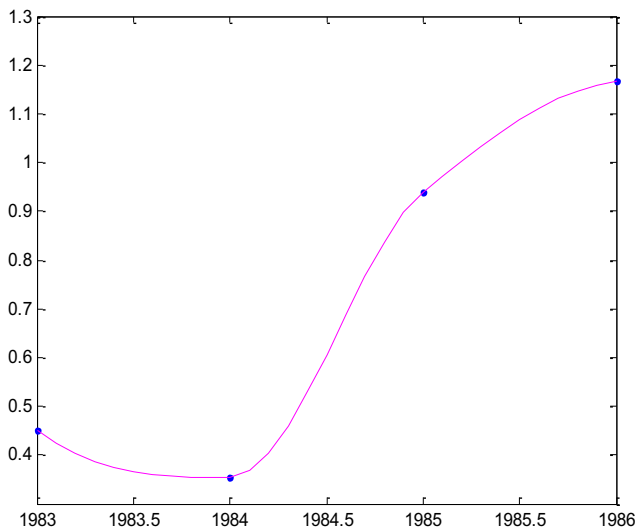
The influenza A broke out in China in 2006. Global influenza pandemic caused by the new influenza A virus in 2009, of which 0.3 million people died (Ren and Gao, 2011; Girard et al., 2010). As shown in Fig. 4, we can observe the composite index degree reached the peak in 2006 and 2009, the composite index degree were slightly higher than the previous year in 2005 and 2008. $I_{2005} - I_{2004} = 0.318385 > 0.08$, $I_{2004} < 1$, $I_{2005} < 1$, $I_{2006} < 1$; $I_{2008} - I_{2007} = 0.322081 > 0.08$, $I_{2007} < 1$, $I_{2008} < 1$, $I_{2009} \geqslant 1$, so the states in 2005 and 2008 are the critical states before the outbreak, the states in 2006 and 2009 are the outbreak states.

The influenza A broke out in India in 2015, of which 1.5 thousand people died (Parida et al., 2016). As shown in Fig. 5, in 2012, the composite index value is 0.63573; 2013 is 0.6060092; 2014 is 0.806321; 2015 is 2.516147. The composite index value was slightly higher than the previous year in 2005 and 2008. So we can presume the influenza a virus reach the critical state before break out in 2005 and 2008, and run up to the outbreak state in 2006 and 2009. $I_{2014} - I_{2013} =$
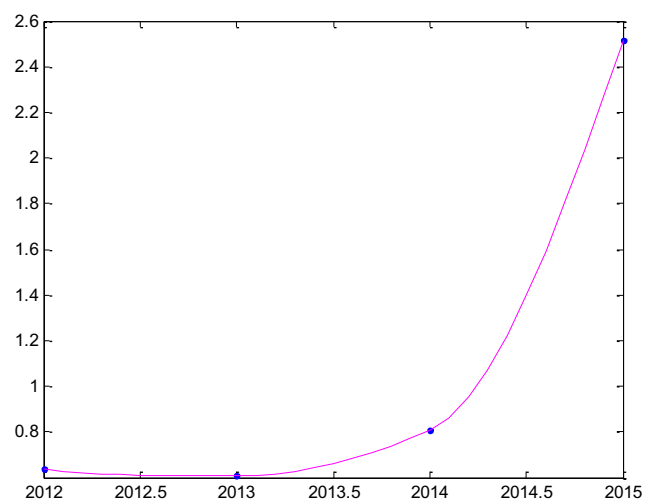

Figure 3 Trend Chart of composite index values from 1983 to 1986.

Figure 5 Trend Chart of composite index values from 2012 to 2015.

$0.273459 > 0.08$, $I_{2013} < 1$, $I_{2014} < 1$, $I_{2015} \geqslant 1$, so the state in 2014 is the critical state before the outbreak, the state in 2015 is the outbreak state.

In general, the composite index varies along with the state of pandemic influenza, which gives a clue showing the turn point of outbreak. The low value (< 1) steady state of the composite index means influenza A is normally in the relatively steady stage. Meanwhile, if the composite index of a certain year increases by more than 0.8 relative to the previous year, and it is less than 1 and it increases sharply and reaches a peak being larger than 1 in next year, it means the year is normally in the critical state before outbreak and the next year is normally in the outbreak state. Therefore, we can predict the outbreak of influenza A and identify the critical state before influenza A outbreak or outbreak state by observing the variation of index value.

5. Conclusion

We select the data of protein amino acid sequence of pandemic influenza virus between 1933 and 2015 in which only some data in a very few years are absent, and obtain a composite index by using the nature of dynamic network biomarkers. The network markers and other traditional markers cannot provide an early warning signal of the critical state before pandemic outbreak in comparison with dynamic network biomarker. Although the amino acid sequence of each protein will randomly fluctuate, the composite index can still provide reliable, significant early warning information when influenza pandemic is close to the critical state or outbreak state. This fully shows the dynamic network biomarker is more stable and accurate to determine the state in which the pandemic influenza virus, particularly the critical state of pandemic influenza. This will achieve the aim of early warning and then strengthen preventive measures in advance. This is of great significance for the research and warning of pandemic influenza virus.

References

Banerjee, R., Roy, A., Das, S., et al, 2015. Similarity of currently circulating H1N1 virus with the 2009 pandemic clone: viability of an imminent pandemic. Infect. Genet. Evol. 32, 107–112.

Chang, L.Y., Shih, S.R., Shao, P.L., et al, 2009. Novel Swine-origin influenza virus A (H1N1): the First Pandemic of the 21st Century. J. Formos. Med. Assoc. 108 (7), 526–532.

Chen, L.N., Liu, R., Liu, Z.P., Li, M.Y., Aihara, K., 2012. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci. Rep. 2, 342–349.

Girard, M.P., Tam, J.S., Assossou, O.M., Kieny, M.P., 2010. The 2009 A (H1N1) influenza virus pandemic: a review. Vaccine 28 (31), 4895–4902.

He, J.L., Hsieh, M.S., Juang, R.H., et al, 2014. A monoclonal antibody recognizes a highly conserved neutralizing epitope on hemagglutinin of H6N1 avian influenza virus. Vet. Microbiol. 174 (3–4), 333–341.

Jin, G.X., Zhou, X.B., Wang, H.H., Zhao, H., Cui, K., Zhang, X.S., Chen, L.N., Hazen, S.L., Li, K., Wong, S.T.C., 2008. The knowledge-integrated network biomarkers discovery for major adverse cardiac events. J. Proteome Res. 7 (9), 4013–4021.

Liu, R., Li, M.Y., Liu, Z.P., Wu, J.R., Chen, L.N., Aihara, K., 2012. Identifying critical transitions and their leading biomolecular networks in complex diseases. Sci. Rep. 2, 813–821.

Liu, R., Wang, X.D., Aihara, K., Chen, L.N., 2014. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers and dynamical network biomarkers. Med. Res. Rev. 34 (3), 455–478.

Ludwig, J.A., Weinstein, J.N., 2005. Biomarkers in cancer staging, prognosis and treatment selection. Nat. Rev. Cancer 5 (11), 845–856.

Parida, M., Dash, P.K., Kumar, J.S., Joshi, G., Tandel, K., Sharma, S., Srivastava, A., Agarwal, A., Saha, A., Sarawat, S., 2016. Emergence of influenza a (h1n1) pdm09 genogroup 6b and drug resistant virus. Euro Surveill. 21 (5), 6–11.

Pan, Y.N., Lou, J.J., Han, X.P., 2014. Outbreak patterns of the novel avian influenza (H7N9). Physica A 401, 265–270.

Ren, D., Gao, J., 2011. Early-warning signals for an outbreak of the influenza pandemic. Chin. Phys. B 20, 1287011–1287014.

Simon, R., 2005. Development and validation of therapeutically relevant multi-gene biomarker classifiers. J. Natl Cancer Inst. 97 (12), 866–867.

Sun, Li., Depuy, G.W., Evanc, G.W., 2014. Multi-objective optimization models for patient allocation during a pandemic influenza outbreak. Comput. Oper. Res. 51, 350–359.

Wu, D.J., Rice, C.M., Wang, X.D., 2012. Cancer bioinformatics: a new approach to systems clinical medicine. BMC Bioinf. 13, 71–80.

Yao, J., Cong, L., Zhu, B.L., Wang, T., 2015. Effect of dietary approaches to stop hypertension diet plan on pregnancy outcome patients with gestational diabetes mellitus. Bangladesh J.Pharmacol. 10 (4), 732–738.

Zhang, H.B., 2016. Element modeling and dynamic property parameters analysis of self-anchored suspension bridge. J. Mech. Eng. Res. Dev. 39 (1), 39–45.