Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Pharmaceutical and Biomedical Analysis 41 (2006) 246–250

JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS

www.elsevier.com/locate/jpba

Short communication

A novel fingerprint map for detecting SARS-CoV

Lei Gao[a], Yong-Sheng Ding[a,1], Hua Dai[a], Shi-Huang Shao[a], Zhen-De Huang[a], Kuo-Chen Chou[a,b,*]

[a] *Bio-Informatics Research Center, College of Information Sciences and Technology, Donghua University, Shanghai 200051, China*

[b] *Gordon Life Science Institute, 13784 Torrey Del Mar, San Diego, CA 92130, USA*

Received 25 May 2005; received in revised form 20 September 2005; accepted 24 September 2005
Available online 14 November 2005

Abstract

Spike (S) protein is the most important membrane protein on the surface of severe acute respiratory syndrome coronavirus (SARS-CoV). It associates with cellular receptors to mediate infection of their target cells. Inspired by such a mechanism, an in-depth investigation into the genome sequences of S protein of SARS-CoV and its receptor are conducted thru a mathematical transformation and graphic approach. As an outcome, a novel method for visualizing the characteristic of SARS-CoV is suggested. An extensive comparison among a large number of genome sequences has proved that the characteristic thus revealed is unique for SARS-CoV. As such, the characteristic can be regarded as the fingerprint map of SARS-CoV for diagnostic usage. Moreover, the conclusion has been further supported in a real case in Guangdong province of China. The fingerprint map proposed here has the merits of clear visibility and reliability that can serve as a complementary clinical tool for detecting SARS-CoV, particularly for the cases where the results obtained by the conventional methods are uncertain or conflicted with each other.

Keywords: SARS-CoV; Fingerprint map; Angiotensin-converting enzyme 2; Spike protein; *Z*-curve; Graphical approach

1. Introduction

Severe acute respiratory syndrome (SARS), which was first recognized in Guangdong province of China in November 2002, emerged as a life-threatening disease associated with pneumonia. A novel coronavirus, called SARS-coronavirus or SARS-CoV, has been proved to be the cause of SARS [1–3]. Since then, progresses in finding inhibitors against SARA from different angles have been reported (see, e.g. [4–9]). It is also important to timely and accurately diagnose SARS-CoV, which can help understand the mechanism of causing the disease and optimize the healing treatment. Unfortunately, it is quite slow to use the routine clinical procedures to diagnose SARS-CoV, as illustrated below.

SARS-CoV-specific RNA can be detected in various clinical specimens such as blood, stool, respiratory secretions, and body tissues by the polymerase chain reaction (PCR) [3]. Although the existing PCR tests were made available during the outbreak, their sensitivity and specificity were unknown because there were no “gold standard” laboratory or clinical definitions for the diagnosis of SARS [10–12]. For instance, the SARS-CoV can be detected by inoculating suitable cell cultures (e.g. Vero cells) with patient specimens (such as respiratory secretions) and propagating the virus in vitro, but the problem is: although the positive cell culture results indicate the existence of SARS-CoV in the sample tested; the negative results do not exclude SARS. Again, although the enzyme-linked immunosorbent assay (ELISA), immunofluorescence assay (IFA), and neutralization test (NT) have the strongpoint of convenience, reliability and repeatability, they are unsuitable for the acute illness [12]. Therefore, the existing diagnostic tests lack sufficient sensitivity for clinical usage in timely ruling out SARS.

From the emergence of SARS-CoV, many studies have been done on its biological medicine aspects, such as pathogeny characteristics, mechanisms of causing disease, clinic diagnosis and treatment, bacterin development and the spreading rules.

Abbreviations: SARS-CoV, severe acute respiratory syndrome coronavirus; ACE2, angiotensin-converting enzyme 2; S protein, spike protein

[*] Corresponding author.

E-mail addresses: ysding@dhu.edu.cn (Y.-S. Ding), kchou@san.rr.com (K.-C. Chou).

[1] To be addressed for request of materials.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.09.031

At the level of molecule biology, studies have been conducted on its genome sequences' characteristic, structures and functions of the translated proteins, and the evolution relationships in sequences. SARS-CoV genomes are distinguished by the presence of a single stranded plus-sense RNA genome about 30 kb in length. Eleven open reading frames (ORF) of the viral genome are translated into 23 proteins, including four main structural proteins: Spike (S) protein, Membrane (M) protein, Envelope (E) protein, and Nucleocapsid (N) protein [13–15]. Research on other known coronaviruses has proved that among the structural proteins of coronaviruses, S protein plays a very important role in virus entry, virus–receptor interactions and their relationship to tropism [16,17]. The S proteins of coronaviruses are large type-I transmembrane glycoproteins that are responsible for receptor building and membrane fusion. On 26 November 2003, angiotensin-converting enzyme 2 (ACE2) was identified as a functional receptor for the SARS-CoV, and S protein associate with cellular receptors to mediate infection of their target cells. Recently, in virtue of giant human antibody libraries, a human monoclonal antibody, which potently neutralizes SARS-CoV and inhibits syncytia formation between cells expressing the S protein and those expressing the SARS-CoV receptor ACE2, was identified [18]. Research on molecular evolution of the SARS-CoV reveals that the highest rate of mutation is seen in the part of the virus genome coding for the S protein. Research data suggests that through the change of S protein, the early SARS-CoV was under significant pressure to mutate in order to become an efficient human virus [19]. In addition, the research on S protein of SARS-CoV will do great favor to clinic diagnosis and bacterin development [20–22].

In view of the important roles of S protein in SARS-CoV's infection and evolution, in the current study we will regard S protein as the primary research object, investigating it at a deeper level. The previous studies on virus gene sequences only concern about the structure or property of the virus gene sequence by itself, seldom relating the investigation to the other objects such as the receptor for the virus. Here we shall take into account the interaction of SARS-CoV and its receptor. It has been observed that that after a certain kind of mathematic transformation, there is an obvious symmetric characteristic shown in gene sequences between SARS-CoV S protein segment and human ACE2; however, no such a characteristic shown in gene sequences between non-SARS-CoV S protein segment and human ACE2. The characteristic can be regarded as a fingerprint map and be applied to SARS virus detection, thus providing a simple and intuitive lab detection method to complement the traditional clinical and epidemiological methods.

2. Methods

Graphical approaches have been successfully used to deal with many biologically interesting problems, such as enzyme kinetics [23–26], protein folding kinetics [27], analysis of codon usage in *E. coli* protein coding sequences [28] and HIV protein sequences [29], among many others. The advantages of graphical approach are in providing an intuitive picture, helping investigators visualize a very complicated or abstract problem and catch its essence or signature (see, e.g. [30–38]). Here let us use graphical approach to study the interaction of SARS-CoV and its receptor ACE2, and see if there is any special feature in the embodiment of their sequences. The mathematic transformation method presented in this paper is based on the graphic approach developed by the previous investigators [28,29,39,40]. Using the graphic approach, or *Z*-curve method [41], we can obtain the transformed graph of the sequences and find the visualization characteristic between SARS-CoV and its receptor ACE2. Also, we do the same analysis on the S protein of non-SARS-CoV.

The *Z*-curve is a novel method for mapping the DNA or RNA sequence into a folding curve in a three-dimensional space. Originally, the DNA (or RNA) sequences are expressed in terms of a series of four letters A, C, G, and $T(U)$ that may be called the letter sequence representation (LSR) of the DNA (or RNA) sequences. *Z*-curve representation is one-to-one correspondence; it is a geometrical approach to express the LSR of DNA (or RNA) sequences. Such a geometric form of DNA (or RNA) sequence displays the new characters of sequence like symmetry, periodicity and local motif, thus providing a brand-new method for decryption of DNA or RNA sequence. The method is presented briefly as follows. Consider one strand DNA or RNA sequence with N bases. Suppose that the cumulative numbers of bases A, C, G, and $T(U)$ occurring in this subsequence from the 1st base to the nth base in the sequence are denoted by A_n, C_n, G_n, and T_n, respectively. The *Z*-curve is composed of a series of nodes P_0, $P_1, P_2, \ldots, P_N$ whose coordinates x_n, y_n and z_n ($n = 0, 1, 2, \ldots, N$) are uniquely determined by the so-called *Z*-transform of DNA (RNA) sequence:

$$\begin{aligned} &x_n = 2(A_n + G_n) - n, \qquad y_n = 2(A_n + C_n) - n, \\ &z_n = 2(A_n + T_n) - n \quad (x_n, y_n, z_n \in [-n, n], \\ &n = 1, 2, 3, \ldots, N) \end{aligned} \tag{1}$$

Let us define $A_0 = C_0 = G_0 = T_0 = 0$. The Z curve is defined as the connection of the nodes P_0, P_1, P_2, $\ldots$, P_N one by one sequentially with straight lines. Define $x_0 = y_0 = z_0 = 0$ so that the *Z*-curve always starts from the origin of the three-dimensional coordinate system.

3. Results and discussion

The gene sequences in our experiment with the annotation information were downloaded from the website of NCBI RefSeq project (http://www.ncbi.nih.gov/RefSeq). We have conducted experiments on the following three kinds of sequences: (1) the S protein segments of more than 80 SARS complete gene sequences; (2) human ACE2 sequences (with the accession numbers of AB046569, NM_021804, AF291820, and AY358714); (3) non-SARS coronavirus sequences including house mouse, cotton aphid, Norway rat (with the accession numbers of BC026801, AB053181, AB053182, YSCACE2, XM_228924, XM_136130, AF502082, and AB122152).

We analyzed the visualization characteristics of the above sequences. The result, derived from the comparisons among a large number of gene sequences, shows a regular quadrilateral

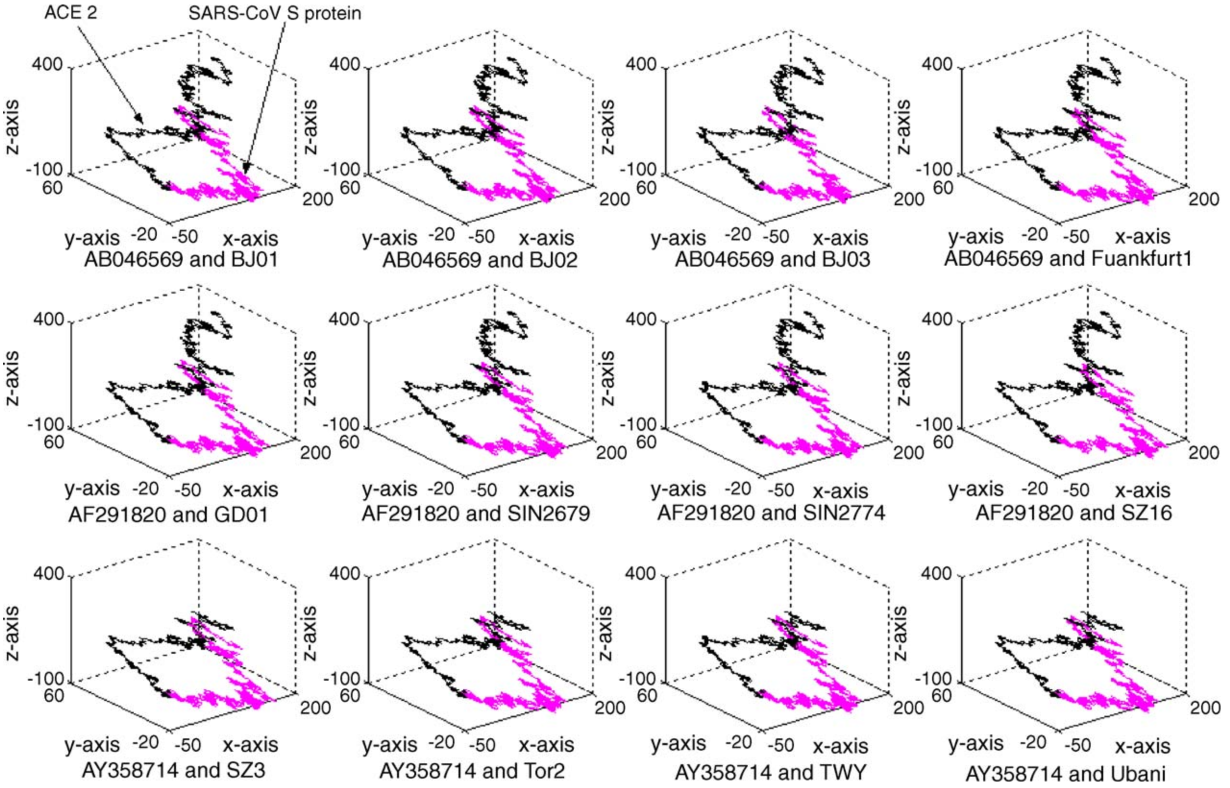

Fig. 1. The fingerprint map of the SARS-CoV S protein segment formed by the mathematic transformation of Eq. (1). In each sub-graph, the black part in the left is formed by human ACE2 gene sequence while the purplish part in the right formed by S protein segment.

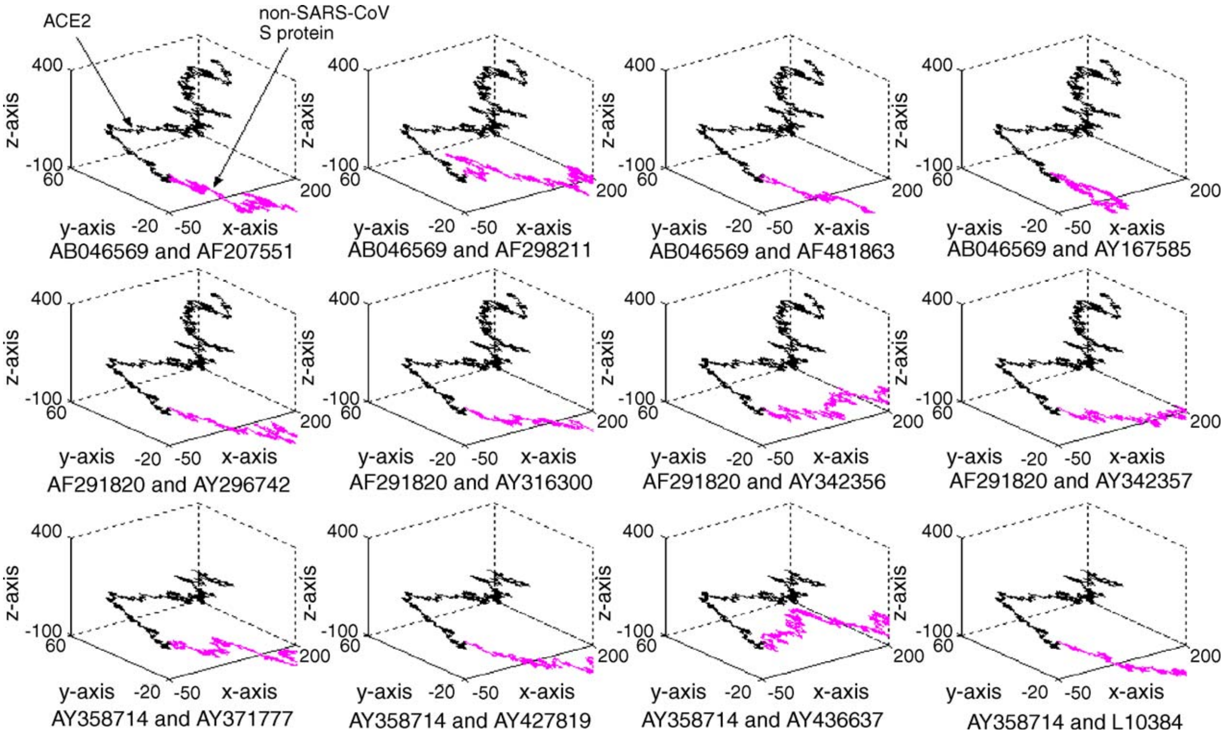

Fig. 2. The representation generated by the same procedure as that of Fig. 1 for the non-SARS coronavirus S protein segment. See the legend of Fig. 1 for further explanation.

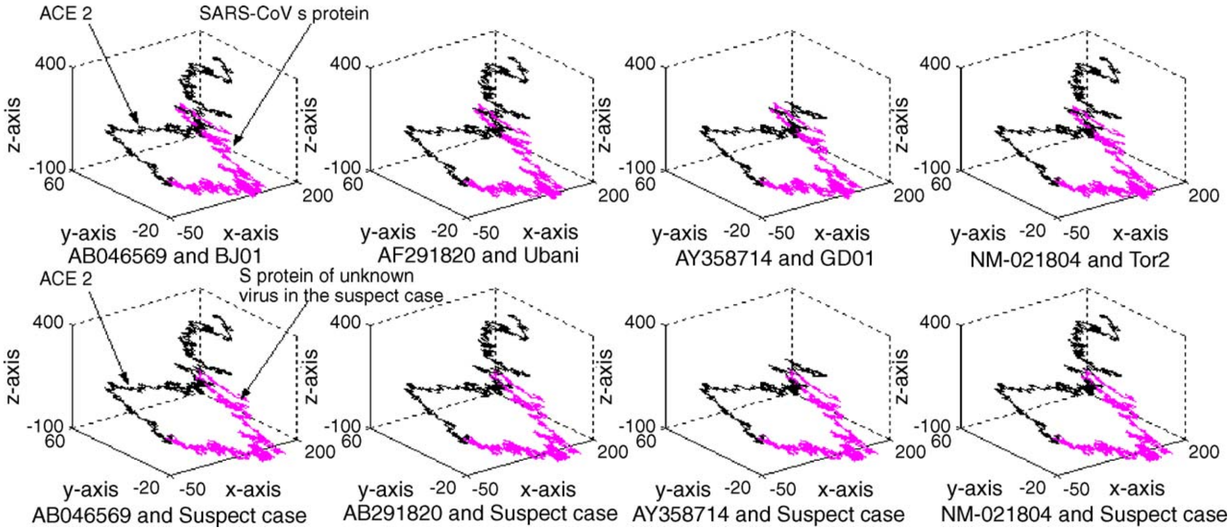

Fig. 3. The representation generated for the corresponding segment from a diagnose SARS case in Guangdong province of China. See the legend of Fig. 1 for further explanation.

between SARS-CoV S protein segment and human ACE2 (cf. Fig. 1). However, there is no such feature between non-SARS coronavirus S protein segment and human ACE2 (cf. Fig. 2). Such a unique feature represents a fingerprint map that can be used to identify and distinguish SARS from other infectious diseases, helping the diagnosis of some ambiguous SARS cases.

Using the fingerprint map, we have analyzed a suspected case in Guangdong province of China. It was quickly found by our method that the S protein of suspected virus did display a regular quadrilateral between the S protein of suspected case and the four kinds of human ACE2 sequences, as shown in Fig. 3. Such character can help us draw the conclusion that this is a SARS-CoV infected case, which has been confirmed by a series of conventional methods performed later.

4. Conclusions

Gene sequence carries all the information of an organism. Since the sequence is a one-dimensional, it is not easy to get the genetic information directly from the sequence itself. The transformation from gene sequence to the visualized graph with a unique feature is an effective way to observe and study the characteristics of gene sequences. Taking the interaction of SARS-CoV and its receptor into consideration, we have obtained a unique characteristic of SARS-CoV using the graphic approach. A visualization interactive relationship has been revealed between SARS-CoV and its receptor human ACE2, which does not exist between the non-SARS-CoV and human ACE2. Such a unique feature can be deemed as the fingerprint map of SARS-CoV that can be used to diagnose suspected cases against SARS. The outcome diagnosed by the fingerprint method has been confirmed by a real SARS case in Guangdong province of China. It has not escaped our notice that the proposed fingerprint map method, along with the method based on cellular automata images reported earlier [42], can also be used to the in-depth research for revealing the molecular mechanism of the disease.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (Nos. 60474037 and 60004006), Program for New Century Excellent Talents in University, and Specialized Research Fund for the Doctoral Program of Higher Education from Educational Committee of China (No. 20030255009).

References

[1] Y.J. Ruan, C.L. Wei, A.L. Ee, V.B. Vega, H. Thoreau, S.T. Su, J.M. Chia, P. Ng, K.P. Chiu, L. Lim, T. Zhang, C.K. Peng, E.O. Lin, N.M. Lee, S.L. Yee, L.F. Ng, R.E. Chee, L.W. Stanton, P.M. Long, E.T. Liu, Lancet 361 (2003) 1779–1785.
[2] J.L. Gerberding, N. Engl. J. Med. 348 (2003) 2030–2031.
[3] C. Drosten, S. Gunther, W. Preiser, S. van der Werf, H.R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A. Fouchier, N. Engl. J. Med. 348 (2003) 1967–1976.
[4] K.C. Chou, D.Q. Wei, W.Z. Zhong, Biochem. Biophys. Res. Commun. 308 (2003) 148–151.
[5] Q.S. Du, S.Q. Wang, D.Q. Wei, Y. Zhu, H. Guo, S. Sirois, K.C. Chou, Peptides 25 (2004) 1857–1864.
[6] S. Sirois, D.Q. Wei, Q.S. Du, K.C. Chou, J. Chem. Inf. Comput. Sci. 44 (2004) 1111–1122.
[7] Q.S. Du, S. Wang, D.Q. Wei, S. Sirois, K.C. Chou, Anal. Biochem. 337 (2005) 262–270.
[8] Q.S. Du, S.Q. Wang, Z.Q. Jiang, W.N. Gao, Y.D. Li, D.Q. Wei, K.C. Chou, Med. Chem. 1 (2005) 209–213.
[9] K.C. Chou, Curr. Med. Chem. 11 (2004) 2105–2134.
[10] K. McIntosh, Clin. Chem. 49 (2003) 845–846.
[11] J.S. Peiris, C.M. Chu, V.C. Cheng, K.S. Chan, I.F. Hung, L.L. Poon, K.I. Law, B.S. Tang, T.Y. Hon, C.S. Chan, K.H. Chan, J.S. Ng, B.J. Zheng, W.L. Ng, R.W. Lai, Y. Guan, K.Y. Yuen, Lancet 361 (2003) 1767–1772.
[12] P. Tang, M. Louie, S.E. Richardson, M. Smieja, A.E. Simor, F. Jamieson, M. Fearon, S.M. Poutanen, T. Mazzulli, R. Tellier, J. Mahony, M. Loeb, A. Petrich, M. Chernesky, A. McGeer, D.E. Low, E. Phillips, S. Jones, N. Bastien, Y. Li, D. Dick, A. Grolla, L. Fernando, T.F. Booth, B. Henry, A.R. Rachlis, L.M. Matukas, D.B. Rose, R. Lovinsky, S. Walmsley, W.L. Gold, S. Krajden, Cmaj 170 (2004) 47–54.

[13] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, N. Engl. J. Med. 348 (2003) 1953–1966.

[14] P.A. Rota, M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M.H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J.L. DeRisi, Q. Chen, D. Wang, D.D. Erdman, T.C. Peret, C. Burns, T.G. Ksiazek, P.E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A.D. Osterhaus, C. Drosten, M.A. Pallansch, L.J. Anderson, W.J. Bellini, Science 300 (2003) 1394–1399.

[15] M.A. Marra, S.J. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S. Butterfield, J. Khattra, J.K. Asano, S.A. Barber, S.Y. Chan, A. Cloutier, S.M. Coughlin, D. Freeman, N. Girn, O.L. Griffith, S.R. Leach, M. Mayo, H. McDonald, S.B. Montgomery, P.K. Pandoh, A.S. Petrescu, A.G. Robertson, J.E. Schein, A. Siddiqui, D.E. Smailus, J.M. Stott, G.S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T.F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G.A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R.C. Brunham, M. Krajden, M. Petric, D.M. Skowronski, C. Upton, R.L. Roper, Science 300 (2003) 1399–1404.

[16] A. Bonavia, B.D. Zelus, D.E. Wentworth, P.J. Talbot, K.V. Holmes, J. Virol. 77 (2003) 2530–2538.

[17] J.J. Breslin, I. Mork, M.K. Smith, L.K. Vogel, E.M. Hemmila, A. Bonavia, P.J. Talbot, H. Sjoestrom, O. Noren, K.V. Holmes, J. Virol. 77 (2003) 4435–4438.

[18] J. Sui, W. Li, A. Murakami, A. Tamin, L.J. Matthews, S.K. Wong, M.J. Moore, A.S. Tallarico, M. Olurinde, H. Choe, L.J. Anderson, W.J. Bellini, M. Farzan, W.A. Marasco, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 2536–2541.

[19] Chinese SARS Molecular Epidemiology Consortium, Science 303 (2004) 1666–1669.

[20] A. Pohl-Koppe, T. Raabe, S.G. Siddell, V. ter Meulen, J. Virol. Met. 55 (1995) 175–183.

[21] C.H. Wang, C.C. Hong, J.C. Seak, Vet. Microbiol. 85 (2002) 333–342.

[22] S.J. Streatfield, J.M. Jilka, E.E. Hood, D.D. Turner, M.R. Bailey, J.M. Mayor, S.L. Woodard, K.K. Beifuss, M.E. Horn, D.E. Delaney, I.R. Tizard, J.A. Howard, Vaccine 19 (2001) 2742–2748.

[23] G.P. Zhou, M.H. Deng, J. Biochem. 222 (1984) 169–176.

[24] K.C. Chou, S.P. Jiang, W.M. Liu, C.H. Fee, Sci. Sin. 22 (1979) 341–358.

[25] K.C. Chou, S. Forsen, J. Biochem. 187 (1980) 829–835.

[26] K.C. Chou, J. Biol. Chem. 264 (1989) 12074–12079.

[27] K.C. Chou, Biophys. Chem. 35 (1990) 1–24.

[28] C.T. Zhang, K.C. Chou, J. Mol. Biol. 238 (1994) 1–8.

[29] K.C. Chou, C.T. Zhang, AIDS Res. Hum. Retroviruses 8 (1992) 1967–1976.

[30] K.C. Chou, F.J. Kezdy, F. Reusser, Anal. Biochem. 221 (1994) 217–230.

[31] I.W. Althaus, A.J. Gonzales, J.J. Chou, M.R. Diebel, K.C. Chou, F.J. Kezdy, D.L. Romero, P.A. Aristoff, W.G. Tarpley, F. Reusser, J. Biol. Chem. 268 (1993) 14875–14880.

[32] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Diebel, K.C. Chou, F.J. Kezdy, D.L. Romero, P.A. Aristoff, W.G. Tarpley, F. Reusser, Biochemistry 32 (1993) 6548–6554.

[33] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Diebel, K.C. Chou, F.J. Kezdy, D.L. Romero, P.A. Aristoff, W.G. Tarpley, F. Reusser, J. Biol. Chem. 268 (1993) 6119–6124.

[34] S.X. Lin, K.E. Neet, J. Biol. Chem. 265 (1990) 9670–9675.

[35] P. Kuzmic, K.Y. Ng, T.D. Heath, Anal. Biochem. 200 (1992) 68–73.

[36] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Diebel, K.C. Chou, F.J. Kezdy, D.L. Romero, P.A. Aristoff, W.G. Tarpley, F. Reusser, Experientia 50 (1994) 23–28.

[37] I.W. Althaus, J.J. Chou, A.J. Gonzales, M.R. Diebel, K.C. Chou, F.J. Kezdy, D.L. Romero, R.C. Thomas, P.A. Aristoff, W.G. Tarpley, F. Reusser, Biochem. Pharmacol. 47 (1994) 2017–2028.

[38] I.W. Althaus, K.C. Chou, K.M. Franks, M.R. Diebel, F.J. Kezdy, D.L. Romero, R.C. Thomas, P.A. Aristoff, W.G. Tarpley, F. Reusser, Biochem. Pharmacol. 51 (1996) 743–750.

[39] C.T. Zhang, K.C. Chou, J. Protein Chem. 12 (1993) 329–335.

[40] C.T. Zhang, K.C. Chou, Amino Acids 10 (1996) 253–262.

[41] L.L. Chen, H.Y. Ou, R. Zhang, C.T. Zhang, Biochem. Biophys. Res. Commun. 307 (2003) 382–388.

[42] M. Wang, J.S. Yao, Z.D. Huang, Z.J. Xu, G.P. Liu, H.Y. Zhao, X.Y. Wang, J. Yang, Y.S. Zhu, K.C. Chou, Med. Chem. 1 (2005) 39–47.