Molecular Mimicry between SARS Coronavirus Spike Protein and Human Protein

Kuo-Yuan Hwa
Department of Molecular Science and Engineering
Institute of Polymeric Science
Center for Biomedical Industries,
National Taipei University of Technology,
Taipei, Taiwan, ROC
Institute of Medical Technology
Taipei Medical University,
Taipei, Taiwan, ROC
e-mail:kyhwa@ntut.edu.tw

Wan Man Lin
Center for Biomedical Industries,
National Taipei University of Technology
Taipei, Taiwan, ROC

Yung-I Hou
Department of Medical Laboratory Science and Biotechnology, College of Medicine, National Cheng Kung University, Taiwan, ROC

Trai-Ming Yeh
Department of Medical Laboratory Science and Biotechnology, College of Medicine, National Cheng Kung University, Taiwan, ROC

Abstract

Molecular mimicry defined as similar structures shared by molecules from dissimilar genes or by their protein products, is a general strategy for pathogens to infect host cells. Therefore, identification of the molecular mimic regions of a pathogen may be helpful to understand the disease. Severe acute respiratory syndrome (SARS) is a new human respiratory infectious disease caused by SARS coronavirus (SARS-CoV). The virus uses the spike (S) protein to interact with the angiotensin converting enzyme 2, the host cell receptor. Our approach is to design a workflow with multiple bioinformatics tools in analyzing the sequence of spike protein of the SARS-CoV in searching its similarity to human proteins. Furthermore, eleven peptides have been synthesized to validate the in silico results.

1. Introduction.

Severe Acute Respiratory Syndrome" (SARS) is a new emerging infectious disease, which was first reported in China in 2002 [1, 2]. The disease is caused by a novel coronavirus (SARS-CoV) [3, 4, 5]. SARS-CoV is a positive-stranded RNA virus. The genome of SARS-CoV is around 29,727 nucleotides in length. The sequence was annotated *in silico* [6]. According to the genomic sequence of SARS-CoV, it is predicted that SARS-CoV produces several structural proteins including spike (S), envelop (E), membrane (M), and nucleocapsid proteins. S protein is a key molecule for the initial viral entry into human cells. The majority of S protein (residues 12-1195) is outside the virus particle, which can be divided into amino-terminal S1 and carboxyl-terminal S2 domain. The S1 domain (residues 12-672) binds to the host cell receptor, angiotensin-converting enzyme 2 (ACE2) while the S2 domain is responsible for membrane fusion [7, 8, 9].

Molecular mimicry, which is defined as similar structures shared by molecules from dissimilar genes or by their protein products, is a general strategy for pathogens to infect host cells and has been proposed as a pathogenic mechanism for autoimmune disease [10]. Therefore, identification of the molecular mimic regions of pathogen may be helpful to understand the disease induced by that pathogen. At present, it is unclear whether molecular mimicry occurs between SARS-CoV S proteins and human peptides. We have approached this question by *in silico* analyzing the sequence of spike protein of SARS-CoV and select regions that share the sequence homology with human proteins. Synthetic peptides were synthesized to validate the prediction.

2. Materials and Methods.

2.1. Peptide prediction and synthesis

Publically available human and coronavirus genome sequences at the National Center for

0-7695-2999-2/07 $25.00 © 2007 IEEE
DOI 10.1109/FBIT.2007.108

Biotechnology Information USA were used for *in silico* prediction. Immunogenic viral peptides were calculated based on the algorithm developed by Kolaskar and Tongaonkar [11]. *In silico* secondary structural analyses of spike protein were performed based on PHD [12] and PREDATOR [13] algorithms. Protein topology prediction was based on the algorithm developed by TMHMM [14]. Hydrophobicity of the peptides was calculated based on the algorithm HMOMENT [15]. Similarity searches between S protein and human genome database were performed by using BLASTP [16]. Extra amino acid residues were added at either N- or C- terminus to decrease the hydrophobicity. Multiple antigen peptides were synthesized by CytoMol Corp (Mountain View, CA, USA).

2.2 Mice immunization

Six to eight-week-old female BALB/c mice were originally purchased from Jackson Laboratory (Bar Harbor, ME, USA) and bred in the Laboratory Animal Center, National Cheng Kung University. Synthetic peptides were emulsified with complete Freund's adjuvant and injected intraperitoneally. Mice were boosted intraperitoneally two weeks after priming. Sera were collected from the axially plexus of the mice at different time intervals.

2.3 SDS-PAGE and Western blot analysis

Proteins in the cell lysate of A549 were separated by 12% SDS-PAGE and transferred to nitrocellulose sheets. Proteins recognized by normal or peptide D08 hyperimmune mice sera were detected by using HRP-conjugated anti-mouse immuno- globulin antibodies and substrates.

3. Results.

3.1 Search for molecular mimic regions in S protein.

The whole amino acid sequence of spike protein was analyzed to find out the potential immunogenic regions and the regions shared sequence homology with human proteins, which is defined as the pathogenic regions. As shown in Figure 1, there are 4 pathogenic regions. Region 1 (residues 199-254), region 2 (residues 658-715), region 3 (residues 893-941), and region 4 (residues 1127-1184) have shared sequence homology with hydroxyacid oxidase, human golgi autoantigen, angrgm-52, and pallidin, respectively. Among these regions, region 3 has the highest score.

Figure 1 *In silico* analysis of S protein.

S protein amino acid sequence was analyzed to find immunogenic regions (yellow regions) and pathogenic regions (regions with shared sequence homology with human proteins, blue regions). Purple regions are both immunogenic and pathogenic regions. Grey region is the leader sequence and brown region is the transmembrane region.

In addition, because des-Arg bradykinin and Ang I are the substrates for ACE2 [17], we also compared the sequence of S protein against bradykinin and Ang I and found that residues 490-502 (**GYQPYR**VVVLSFEE) of S protein showed sequence homology with bradykinin (Figure 2).

Figure 2 Sequence homology between S protein and bradykinin.

Pair wise alignment was calculated based on Smith-Water local alignment with matrix set at EBLOSUM 45. "|"is annotated for identical residues; ":" and "." are for similar residues.

3.2 *In silico* predicted pathogenic peptides.

Eleven antigenic peptides (Table 1), which represent those pathogenic regions were predicted and synthesized.

Table 1. Sequence of the eleven synthetic peptides.

* Extra amino acid residues which were indicated by italic letters were added at either N- or C-terminus to decrease the hydrophobicity.

Peptide	Amino acid positions	Amino acid sequence	No. of amino acids
D01	199-210	**GYQPIDVVRDL***G*	12
D02	658-669	**ASYHTVSLLRST***SQK*	15
D03	733-744	*EEG***NLLLQYGSFCTQ**	15
D04	745-753	*EE***LNRALSGIAGQ**	13
D05	763-770	**VFAQVKQM**	8
D06	911-919	**KAISQIQES***LTTE*	13
D07	927-937	*G***LGKLQDVVNQN***GE*	14
D08	942-951	*A***LNTLVKQLSSN**	12
D09	1154-1162	**INASVVNIQ***K*	10
D10	490-502	**GYQPYRVVVLSFEE**	14
D11	306-317	*G***FRVVPSGDVVR***F*	13

3.3 Hyperimmune sera against D08 cross-reacted with A549 cells

To test whether synthetic peptides D01, D07 D08 can induce antibodies cross-reacted with human proteins, we immunized mice with these peptides to generate hyperimmune sear against these peptides. Using Western blot analysis, hyperimmune sera against D08 could recognize more bands in A549 cell lysate as compared to normal mice sera (Figure 3). In addition, hyperimmune sera against D07 but not D01 showed similar cross-reactivity to A549 cells as hyperimmune sera against D08 did (data not shown).

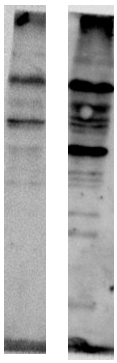

Figure 3 Anti-D08 peptide sera recognizes protein from human A549 cell line.

3.4 Hyperimmune sera against D10 cross-reacted with bradykinin

To test whether the synthetic peptides D10 indeed can induce antibodies cross-reactive with bradykinin and Ang I, as predicted, we immunized mice with D10 peptides to generate hyperimmune sera against this peptide. Significant increase of antibodies against D10 was found in D10 hyperimmune sera, which could cross-react with bradykinin- but not with Ang I-coated plates (Figure 4).

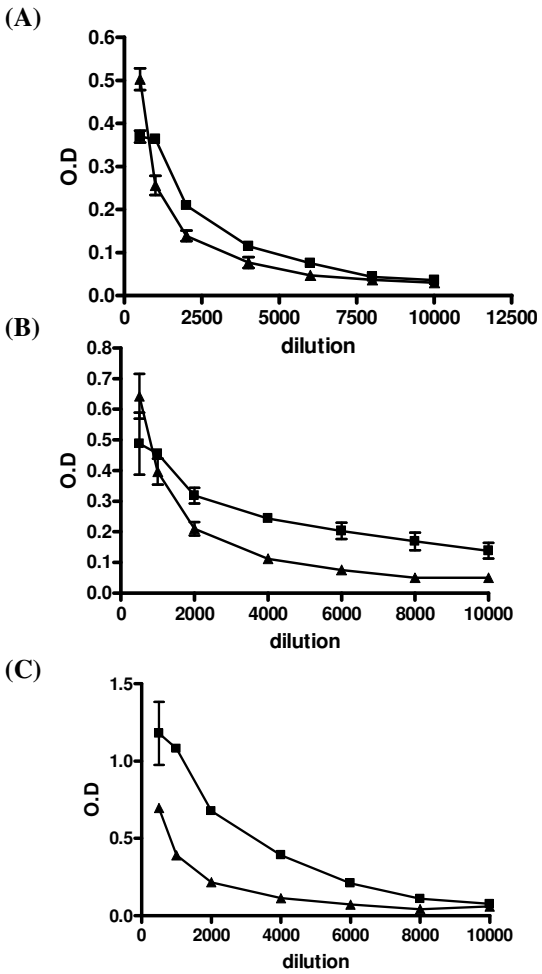

Figure 4 The cross-reactivity of D10 antibody with bradykinin and Ang I.

Hyperimmune sera from D10 immunized mice (■) or normal mice sera (▲) were diluted as indicated and reacted with D10-, bradykinin- or Ang I-coated ELISA plates as indicated. Bound antibodies were detected as described in Materials and Methods. Data represents the mean ± SD of triplicates.

4. Discussion.

In this study we have identified four pathogenic regions of SARS-CoV S protein which share sequence homology with different human proteins. Among them, pathogenic region 3 (residues 893-941), which shares sequence homology with angrgm-52 (GenBank

accession no. AAL62340), a novel gene up-regulated in human mesangial cells stimulated by angiotensin II, may deserve further investigation. Peptides D07 and D08 of this region were recognized by the sera of SARS patient indicating that this region is immunogenic and can be recognized by the immune system during SARS-CoV infection. Murine hyperimmune sera against peptides D07 or D08 were able to bind to recombinant S2 but not S1 domain of S protein (data not shown). In addition, hyperimmune sera against D07 or D08 also bound to the cytoplasmic region of A549 cells and recognized several proteins in the A549 cell lysate. These results indicate that regions represented by D07 and D08 are immunogenic and may induce autoantibodies. However, further study is required to understand the biological function of these regions and the role of their antibodies in the pathogenesis of SARS-CoV infection.

In addition to D07 and D08 peptides, we also noticed D10 peptide which represents residues 490-502 of S1 domain contained some interesting activities. The D10 peptide, which shared sequence homology with bradykinin, was able to generate antibodies cross-reactive with bradykinin. In addition, D10 peptide could stimulate A549 to produce IL-8 and proliferation as Ang I did. These results suggest that the region of D10 in S protein may bind to Ang I receptor, ACE2, and may be involved in the binding of SARS-CoV to ACE2. This is consistent with the previous report, which indicates that residues 318-510 of S1 domain can bind to ACE2 [18] and is similar to the receptor binding domain of the HCoV-229E, which is within a fragment containing residues 407 to 547 [19]. Therefore, region 490-502 of S1 domain may be involved in the receptor binding domain of SARS-CoV.

In summary, our results suggest that molecular mimicry occurs between SARS-CoV and host proteins. Motifs shared sequence homology with host proteins of SARS-COV may be involved in the binding and fusion of SARS-CoV to host cells. Antibody against these motifs may contain neutralization activity against SARS-CoV infection or participate in the immunopathogenesis induced by SARS-CoV. Small molecules derived from molecular structure of these motifs may provide alternative approaches to disrupt the infection of SARS-CoV [20].

5. Acknowledgments

This work was supported by the grant NSC92-2751-B006-Y from the National Science Council, Taipei, Taiwan.

6. Reference

[1] M.D. Christian, S.M. Poutanen, M.R. Loutfy, M.P. Muller, D.E. Low, "Severe acute respiratory syndrome", *Clin Infect Dis,* vol. 38, (2004), pp1420-1427.

[2] T. Kuiken, R.A. Fouchier, M. Schutten, G.F. Rimmelzwaan, G. van Amerongen, D. van Riel, J.D. Laman, T. de Jong, G. van Doornum, W. Lim, A.E. Ling, P.K. Chan, J.S. Tam, M.C. Zambon, R. Gopal, C. Drosten, S. van der Werf, N. Escriou, J.C. Manuguerra, K. Stohr, J.S. Peiris, A.D. Osterhaus, "Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome", *Lancet,* vol. 362, (2003), pp 263-270.

[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, A. Berger, A.M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J.C. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H.D. Klenk, A.D. Osterhaus, H. Schmitz, H.W. Doerr, "Identification of a novel coronavirus in patients with severe acute respiratory syndrome", *N Engl J Med*, vol . 348, (2003), pp1967-1976.

[4] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, P.E. Rollin, S.F. Dowell, A.E. Ling, C.D. Humphrey, W.J. Shieh, J. Guarner, C.D. Paddock, P. Rota, B. Fields, J. DeRisi, J.Y. Yang, N. Cox, J.M. Hughes, J.W. LeDuc, W.J. Bellini, L.J. Anderson, "A novel coronavirus associated with severe acute respiratory syndrome", *N Engl J Med*, vol. 348, (2003), pp1953-1966.

[5] 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-Rassmussen, R. Fouchier, S. Gunther, A. D. Osterhaus, C. Drosten, M.A. Pallansch, L.J. Anderson, W.J. Bellini, "Characterization of a novel coronavirus associated with severe acute respiratory syndrome", *Science*, vol. 300, (2003), pp1394-1399.

[6] 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, "The genome sequence of the SARS-associated coronavirus", *Science*, vol. 300, (2003), pp1399-1404.

[7] Y. He, Y. Zhou, H. Wu, B. Luo, J. Chen, W. Li, S. Jiang, "Identification of Immunodominant sites on the spike protein of severe acute respiratory syndrome (SARS) coronavirus:

implication for developing SARS diagnostics and vaccines", *J Immunol*, vol. 173, (2004), pp4050-4057.
[8] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A.Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, H. Choe, M. Farzan, "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus", *Nature*, vol. 426 (2003), pp450-454.
[9] S.K. Wong, W. Li, M.J. Moore, H. Choe, M. Farzan, "A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2", *J Biol Chem*, vol 279, (2004), pp3197-3201.
[10] M.B. Oldstone, "Molecular mimicry and immune-mediated diseases", *FASEB J* , vol 12, (1998), pp1255-1265.
[11] A.S. Kolaskar, P.C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens", *FEBS Lett*, vol 276, (1990), pp172-174.
[12] B. Rost, G Yachdav and J. Liu, "The PredictProtein Server", *Nucleic Acids Research* (web service issue), 32, (2004), pp321-326
[13] D. Frishman and P. Argos, "Incorporation of long-distance interactions into a secondary structure prediction algorithm. Protein Engineering", vol. 9, (1996) , pp133-142.
[14] A. Krogh, B. Larsson, G. von Heijne, and E.L.L. Sonnhammer., "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes", *Journal of Molecular Biology*, vol. 305, (2001), pp567-580,.
[15] D. Eisenberg, R.M. Weiss, T.C. Terwilliger, "The hydrophobic moment detects periodicity in protein hydrophobicity." *Proc Natl Acad Sci U S A*, vol 81, (1984), pp140-4.
[16] W. Gish, D. J. States, "Identification of protein coding regions by database similarity search.", *Nature Genet.*, vol. 3, (1993) pp266-272.
[17] M. Donoghue, F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, R. Jeyaseelan, R. E. Breitbart, S. Acton, "A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9", *Circ Res*, vol. 87, (2000), pp1-9.
[18] S.K. Wong, W. Li, M.J. Moore, H. Choe, M. Farzan, A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2, *J Biol Chem*, vol. 279, (2004), pp3197-3201.
[19] J.J. Breslin, I. Mork, M.K. Smith, L.K. Vogel, E.M. Hemmila, A. Bonavia, P.J. Talbot, H. Sjostrom, O. Noren, KV. Holmes, "Human coronavirus 229E: receptor binding domain and neutralization by soluble receptor at 37 degrees C", *J Virol*, vol. 77, (2003), pp4435-4438.
[20] R.Y. Kao, W.H. Tsui, T.S. Lee, J.A. Tanner, R.M. Watt, J.D. Huang, L. Hu, G. Chen, Z. Chen, L. Zhang, T. He, K.H. Chan, H. Tse, A.P. To, L.W. Ng, B.C. Wong, H.W. Tsoi, D. Yang, D.D. Ho, K.Y. Yuen, "Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetic", *J Chem Biol*, vol. 11, (2004), pp1293-1299.