RESEARCH NOTES

Bioinformatics analysis of SARS-Cov M protein provides information for vaccine development*

LIU Wanli, LU Yun and CHEN Yinghua**

(Laboratory of Immunology, Research Centre for Medical Science and Department of Biology, Tsinghua University; Protein Science Laboratory of MOE, Beijing 100084, China)

Received July 17, 2003; revised August 11, 2003

Abstract The pathogen causing severe acute respiratory syndrome (SARS) is identified to be SARS-Cov. It is urgent to know more about SARS-Cov for developing an efficient SARS vaccine to prevent this epidemic disease. In this report, the homology of SARS-Cov M protein to other members of coronavirus is illustrated, and all amino acid changes in both S and M proteins among all available SARS-Cov isolates in GenBank are described. Furthermore, one topological trans-membrane secondary structure model of M protein is proposed, which is corresponded well with the accepted topology model of M proteins of other members of coronavirus. Hydrophilic profile analysis indicated that one region (aa150~210) on the cytoplasmic domain is fairly hydrophilic, suggesting its property of antigenicity. Based on the fact that cytoplasmic domain of the M protein of some other coronavirus could induce protective activities against virus infection, this region might be one potential target for SARS vaccine development.

Keywords: SARS-Cov, M protein, anti-SARS vaccine strategy.

Several months ago, an outbreak of severe acute respiratory syndrome (SARS) started to spread around the world. As of this writing (May 14, 2003), more than 7500 persons have been infected in as many as 29 countries[1]. The pathogen causing SARS has already been identified to be SARS-Cov by a number of laboratories worldwide[2,3]. Severe diseases in animals, such as the diseases caused by porcine transmissible gastroenteritis virus (TGEV), murine hepatitis virus (MHV) and porcine hemagglutinating encephalomyelitis virus (PHEV), have ascribed to some coronavirus members. However, only mild diseases, such as infection of the lower respiratory tract and necrotizing enterocolitis, were related with two known human coronavirus (HcoVs), HcoVs-229E and HcoVs-OC43. So SARS-Cov might be the first human coronavirus causing severe disease in humans concerning its severe clinical syndrome and mortality rate of 3%~10%[4].

Two major surface proteins, S (Spike) protein and M (membrane) protein, might be contained in SARS-Cov. S protein is involved in virus attachment on target cell receptors and virus cell fusion. M protein, embedded in the membrane of the virus particles, is associated with the RNA and nucleocapsid protein complex by its cytoplasmic domain[5]. In this study, we analyzed the homology of both SARS-Cov S and M proteins with other coronavirus members, and identified all of the amino acid changes in both proteins among all available sequences of SARS-Cov isolates in GenBank[6]. Besides, one topological trans-membrane secondary structure model of M protein was proposed based on the results obtained from using two trans-membrane domain prediction programs, TMHMM and TMpred, and one protein secondary structure prediction program, PROFsec[7,8]. Finally, one possible target on cytoplasmic domain of M protein was proposed for SARS-Cov vaccine development according to the immunogenicity of other coronavirus M proteins.

Protein homology analysis showed that SARS-Cov M protein was 61% homogenous to the other two PHEV strains (strain 67N, IAF-404), while the S protein was most related to three MHV strains (strain Wb1, JHM and A59) with a homology of 45% (Table 1). Alignment of the amino acid se-

* Supported by the Anti-SARS Fund of Tsinghua University and the Ministry of Science and Technology, and the National Found for Distinguished Young Scholars (Grant No. 30025038)

** To whom correspondence should be addressed. E-mail: chenyh@mail.tsinghua.edu.cn

quences showed three mutations in S protein and one mutation in M protein among all five SARS-Cov isolates in GenBank (Table 2). In detail, all three amino acid mutations in S protein, aa78 (Asp to Gly), aa245 (Thr to Ile) and aa578 (Ser to Ala), belong to amino acid substitution with quite different R groups: aa78 (Asp to Gly), Asp has a negatively charged R group, while the substituted Gly has a nonpolar and aliphatic R group; aa245 (Thr to Ile) and aa578 (Ser to Ala), Thr and Ser contain polar and uncharged R groups, while Ile and Ala contain nonpolar and aliphatic R groups. Interestingly, the only mutation in M protein belongs to amino acid substitution with similar R groups: aa155 (Ser to Pro), both of Ser and Pro have polar and uncharged R group. These results proved that more mutation sites and more possible alteration of chemical or physical features happened in S protein than M protein, which might induce possible changes of immunogenicity. Although both S and M are all directly related with inducing protective activity, and more sights are focusing on S protein, M protein should be deserved to be one competitive vaccine protein concerning the above positive characters.

Table 1. Homology of SARS-Cov S and M proteins to other members of coronavirus

SARS-Cov BJ01	Coronavirus strain (Strain, GenBank accession No.)	Homology rate (%) (Score)
S protein (1255 aa)	MHV (Wb1, AAB30950)	45 (495)
	MHV (JHM, CAA28484)	45 (494)
	MHV (A59, AF208066)	45 (483)
M protein (221 aa)	PHEV (67N, AAL80035)	61 (167)
	PHEV (IAF-404, AAM77004)	61 (166)
	BCov (ENT, AAK83361)	60 (163)
	MHV (TY, AAF05705)	60 (162)

Table 2. Amino acid changes in S and M proteins among all SARS-Cov isolates in GenBank

S protein (1255 aa)				M protein (221 aa)	
SARS-Cov isolates	aa78	aa245	aa578	SARS-Cov isolates	aa155
BJ01 (AY278488)	Asp	Thr	Ser	BJ01 (AY278488)	Ser
Tor2 (NC_004718)	Gly	Ile	Ala	Tor2 (NC_004718)	Ser
Urbani (AY278741)	Gly	Ile	Ser	Urbani (AY278741)	Pro
CUHK-W1 (AY278554)	Asp	Thr	Ser	CUHK-W1 (AY278554)	Ser
UHK-Su10 (AY282752)	Gly	Thr	Ser	BJ03 (AY278490)	Ser

As one RNA viruses, SARS-Cov has the potential to mutate to escape host defenses. Developing vaccine with multi components, which is composed of several protective epitopes based on both S protein and M protein, should be one promising strategy against SARS-Cov mutation according to our previous work on HIV and influenza virus[9]. Especially, Anton found that neutralizing TGEV-specific antibodies were induced when porcine TGEV-immune cells were stimulated with S+M combination, but no such antibodies were induced when stimulated with either S or M protein alone. What is more, $>60\%$ of neutralizing antibodies were induced in S and M protein combined group compared with the whole TGEV virions group[10]. These positive descriptions should also induce more interests of M protein in anti-SARS vaccine development besides the S protein.

Most coronavirus M protein spans the viral membrane bilayer three or four times, leaving a short N-terminal domain outside of the virus and a long C-terminal inside[11]. Here, we predicted the possible trans-membrane domains of SARS-Cov M protein using TMHMM and TMpred analysis server. As a result, three trans-membrane helices, located approximately at aa21 ~ 38, 51 ~ 69 and 74 ~ 99, might be contained in M protein. 21 N-terminal amino acid sequences were outside of the virus particle, and 121 C-terminal amino acid residues inside. Furthermore, we predicted the secondary structure of SARS-Cov M protein by PROF_sec server. Corresponded well with the trans-membrane domain prediction results, helices are recommended at aa13 ~ 103, where three trans-membrane helices domains are predicted formerly. Several sheets alternated by short loops are recommended at aa110 ~ 221, which is exactly the former predicted cytoplasmic domain. So the M protein of SARS-Cov seems to have similar spatial structure with other coronavirus members, and we tried to describe the topological trans-membrane secondary structure model of M protein in Fig. 1, which integrated the prediction results of both secondary structure and the trans-membrane domain prediction results.

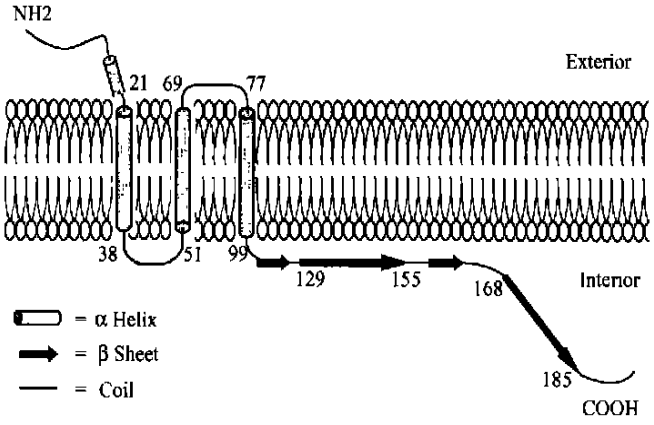

Fig. 1. Possible topological trans-membrane secondary structure model of SARS-Cov M protein. Exterior means outside of virions, and interior mean inside of virions.

Protective epitopes on SARS-Cov should be found out first to obtain efficient vaccines to control SARS epidemic in future. Some former studies have proved that the M protein of some coronavirus members could induce protective activities. Two epitopes were found in the cytoplasmic domain of MHV M protein, and monoclonal antibodies against cytoplasmic domain of TGEV M protein could neutralize virus[12, 13]. Besides, the hydrophilic domain of membrane protein on influenza virus, but not the trans-membrane domain, was found to induce antibodies with efficient and broad-spectrum antiviral activities by others and us[14]. To know further M protein character for vaccine development, hydrophilic profile was drawn based on hydropathy index of all 221 amino acid residues of SARS-Cov M protein[15]. The hydrophobic sequence is mainly at the former predicted trans-membrane domain, while most parts of cytoplasmic domains are hydrophilic, which correspond well with the topological trans-membrane domain model. More importantly, the region (aa150 ~ 210) on M protein should be existed in the aqueous environment, which is more convenient for humoral immune response compared with other hydrophobic regions on M protein (Fig. 2). Although these analyses are very preliminary, they imply this region might be one potential target sequence for anti-SARS vaccine strategy. Besides, our previous work suggested a putative receptor-binding region (aa51 ~ 54, aa195 ~ 197) on SARS-Cov S protein[16]. Multi epitope vaccine, which combines these putative target sequences on both S proteins and M proteins of SARS-Cov, would be one interesting anti-SARS vaccine development strategy and deserves further research.

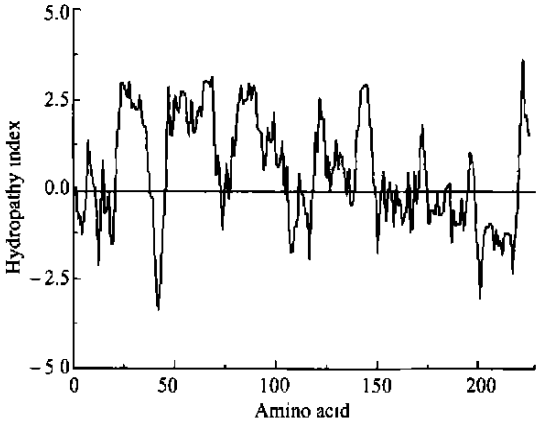

Fig. 2. Hydrophilic profile of all 221 amino acid residues of SARS-Cov M protein. Hydropathy index is usually used to evaluate whether one certain amino acid will be found in an aqueous environment ("−" values) or in a hydrophobic environment ("+" values).

References

1 Cumulative number of reported cases of severe acute respiratory syndrome (SARS). WHO, update 54 2003: http://www.who.int/csr/sars/archive/2003_05_13/en/ (Accessed May 14, 2003).
2 Drosten C. et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med., 2003, 348: 1967.
3 Peiris, J. S. M. et al. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet, 2003, 361: 1319.
4 Marco, A. M. et al. The genome sequence of the SARS-associated coronavirus. Science, 2003, 300: 1399.
5 Ruan, Y. J. et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet, 2003, 361: 1779.
6 Stephen, A. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 1997, 25: 3389.
7 Krogh, A. et al. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol., 2001, 305: 567.

8 Rost, B. et al. Redefining the goals of protein secondary structure prediction. J. Mol. Biol., 1994, 235: 13.
9 Xiao, Y. et al. Epitope-vaccine as a new strategy against HIV-1 mutation. Immunol Letters, 2001, 77: 3.
10 Anton, I. M. et al. Cooperation between transmissible gastroenteritis coronavirus (TGEV) structural proteins in the *in vitro* induction of virus-specific antibodies. Virus Res., 1996, 46: 111.
11 Gabriella, E. et al. Recombinant M protein-based ELISA test for detection of antibodies to canine coronavirus. J. Virol. Methods, 2003, 109: 139.
12 Tooze, S.A. et al. Identification of two epitopes in the carboxy-terminal 15 amino acids of the E1 glycoprotein of mouse hepatitis virus A59 by using hybrid proteins. J. Virol., 1986, 60: 928.
13 Risco, C. et al. Membrane protein molecules of transmissible gastroenteritis coronavirus also expose the carboxy-terminal region on the external surface of the virion. J. Virol., 1995, 69: 5269.
14 Liu, W. L. et al. N-terminu of M2 protein could induce antibodies with inhibitory activity against influenza virus replication. FEMS Immunol Med. Mic., 2003, 35: 141.
15 Kyte, J. et al. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol.. 1982, 157: 105.
16 Lu, Y. et al. Spike protein homology between the SARS-associated virus and murine hepatitis virus implies existence of a putative receptor-binding region. Chinese Science Bulletin, 2003, 48(11): 1115.