VIRAL IMMUNOLOGY
Volume 17, Number 4, 2004

Pp. 528–534

Structure-Based Preliminary Analysis of Immunity and Virulence of SARS Coronavirus

YAN LI,[1,2]* CHUNQING LUO,[1]* WEI LI,[1] ZHAO XU,[1] CHANGQING ZENG,[1] SHENLI BI,[3] JUN YU,[1] JUN WU,[2] and HUANMING YANG[1]

ABSTRACT

The research on SARS-associated coronavirus (SARS-CoV) has not stopped since its discovery, but the pathogenesis of SARS is still unclear. To explore the possible molecular mechanisms of the invasion and virulence of SARS-CoV, we investigated the structural basis of the viral proteins using computational biology. Forty-five motifs relating to superantigens, toxins and other bioactive molecules were detected in the proteins of SARS-CoV. The results showed that the distribution of the motifs varied in different proteins. Enzyme-like motifs were located in the R protein, while ICAM-1–like and toxin-like molecules were located in the spike, envelop, nucleocapsid, PUP1, PUP 2 and PUP 4 proteins. Comparison of SARS-CoV with other viruses (OC43, PEDV, HRSV, HHerpV and HAdenoV) showed that each group of motifs was different for each type of virus. Data suggest that the proteins of SARS-CoV with toxic motifs might play crucial roles in targeting host cells and interfering with the immune system. This study provides new information for drug and vaccine design, as well as therapeutic strategies against SARS.

INTRODUCTION

THE SEVERE ACUTE RESPIRATORY SYNDROME (SARS), with high rates of morbidity and mortality, has affected thousands of people and killed hundreds of them since November of 2002. The SARS-associated coronavirus (SARS-CoV) was first identified as the pathogen of this disease in April 2003 (3,6).

The most common symptoms of the disease were fever (>38.5°C), dry cough, myalgia, short breath, and dyspnea. The progress of the illness was very fast. Conventional experimental tests show lymphopenia and slight leukopenia. Furthermore, traditional anti-bacteria treatment has little effect on this disease (15).

Based on the chest radiographs and histopathological investigation, the pathological changes were characterized by massive infiltration and marked alveolar edema with hemorrhage and hyaline membrane formation, even atrophy of lymph and widely angitis, but few inflammatory cells were observed. Furthermore, the serological evidence showed an increased level of IgG at the second week after onset of the symptoms (12,15). But the results of flow cytometry demonstrated both $CD4^+$ and $CD8^+$ T cells significantly decreased and would recover as soon as the symptoms disappeared (7). The above clinical evidence suggested that viral damage and allergic immune response could play a major role in the illness and may finally result in the Adult Respiratory Distress Syndrome (ARDS), which could be lethal to the patients.

There is no doubt that the special proteins of SARS-

[1]Beijing Genomics Institute (BGI), Chinese Academy of Sciences (CAS), Beijing, China.
[2]The Third Military Medical University, Chongqing, China.
[3]Center of Disease Control and Prevention, Beijing, China.
*These authors contributed equally to the work.

CoV determine its invasion and virulence against its host (especially human). In a previous study (13), we have identified four major structural proteins, namely, the spike protein (S protein), the envelop protein (E protein), the membrane protein (M protein), and the nucleocapsid protein (N protein), as well as five putative uncharacterized proteins (PUPs), which are all exogenous substances to the human body. Now, we investigate those motifs of SARS-CoV in regard to the immunity and toxicity in order to reveal the possible molecular mechanism of the disease.

MATERIALS AND METHODS

Source of sequences. The SARS-CoV BJ01 isolate was selected for the analysis, the genome of which had been sequenced in our previous study (13). The putative proteins encoded in the genome were analyzed. We also analyzed the proteins encoded in the genomes of the human coronavirus OC43 (OC43, NC_005147), the porcine epidemic diarrhea virus (PEDV, NC_003436), the human respiratory syncytial virus (HRSV, NC_001781), the human adenovirus A (HAdenoV, NC_001460) and the human herpesvirus (HHerpV, NC_001806). At the same time, a local database was set up on an IBM p690 machine, totally containing 62,228 known sequences of protein molecules from the public database at NCBI, of which 38,010 sequences were associated with the antigens, 861 sequences with superantigens (sAgs), 4,773 sequences with cytokines, and 18,584 sequences with toxins. Because the majority of these sequences were derived from experimental study on the functions of the protein, this database was used to detect the functional motifs by sequence alignment.

Sequence analysis. The protein sequences of SARS-CoV were aligned using BLAST (ftp://ftp.ncbi.nih.gov/blast/) against the local database as described above. The blast results with identity value more than 30% were statistically analyzed. Redundant data and those sequences less than 50 amino acids showing less than 35% identity were removed. Selected segments were classified according to the annotations. NetNGlyc1.0 software was used to detect glycosylation sites of the proteins (www.cbs.dtu.dk/services/NetNGlyc). The physical and chemical features of each peptide were examined by using Compute pI/MW, ProtScale (http://us.expasy.ch/tools) and Genhan (www.genhan.net/ peptide1.htm).

Assess the conservation of the motifs. The conservation of the motifs which were less than 50 amino acids showing 35–39% identity, were compared and examined against other six types of coronaviruses, including the avian infectious bronchitis virus, the human coronavirus 229E, the porcine epidemic diarrhea virus, the human coronavirus OC43 and the murine hepatitis virus.

RESULTS

Distribution of the identified motifs in the SARS-CoV. Overall, 45 entries of sequences related to antigenicity and toxicity were identified, by comparing protein sequences of the SARS-CoV with the local database containing 62,228 known sequences. The average length of the identified motifs was 39.5 ± 19.4 residue, and the average identity was 40%.

The motifs which were less than 50 amino acids showing 35–39% identity, were examined for their conservation in six coronaviruses. The data showed that all of them could be found in more than two other coronaviruses with identities from 43% to 75%. The average identity is 51.7%, in contrast to 43.9% of the globe alignments against the whole amino acid sequences. It suggested that all the motifs have relatively higher conservation in the family of coronavirus.

The motifs were distributed over nine ORFs of viral nonstructural and structural proteins, and the different coverage in each protein was shown in Table 1. In overview, the distribution was sparser in the long R protein sequence than in others, but it was denser in the short PUP1, PUP4 and PUP5.

The R protein of SARS-CoV is a type of polyprotein encoded by the biggest ORF accounting for almost two thirds of the whole viral genome. Of 15 identified motifs in the R protein, six were associated with enzyme, and in agreement with their postulated functions, three were likely to show antigenicity, and the other three were associated with toxin in this biggest R protein (Fig. 1). Actually the R protein is not present in the virion, therefore the possible antigenicity or virulence should be related to post-translational modification, such as cleavage by one of the viral protease (16). In the leader protein, a subunit near the N-terminal of R protein, one motif was likely to be related to the plasminogen activation.

Nine motifs localized in the S protein had a relatively even distribution along the entire amino acid sequence.

TABLE 1. COVERAGE OF THE PREDICTED MOTIFS IN EACH PROTEIN OF SARS-COV

Protein	*Length (aa)*	*Number*	*Coverage (%)*
R protein	7073	15	8.91%
S protein	1255	9	27.65%
E protein	76	2	44.74%
M protein	221	2	30.77%
N protein	422	4	41.47%
PUP1	274	4	63.14%
PUP2	154	2	31.17%
PUP4	122	5	94.26%
PUP5	98	2	77.55%

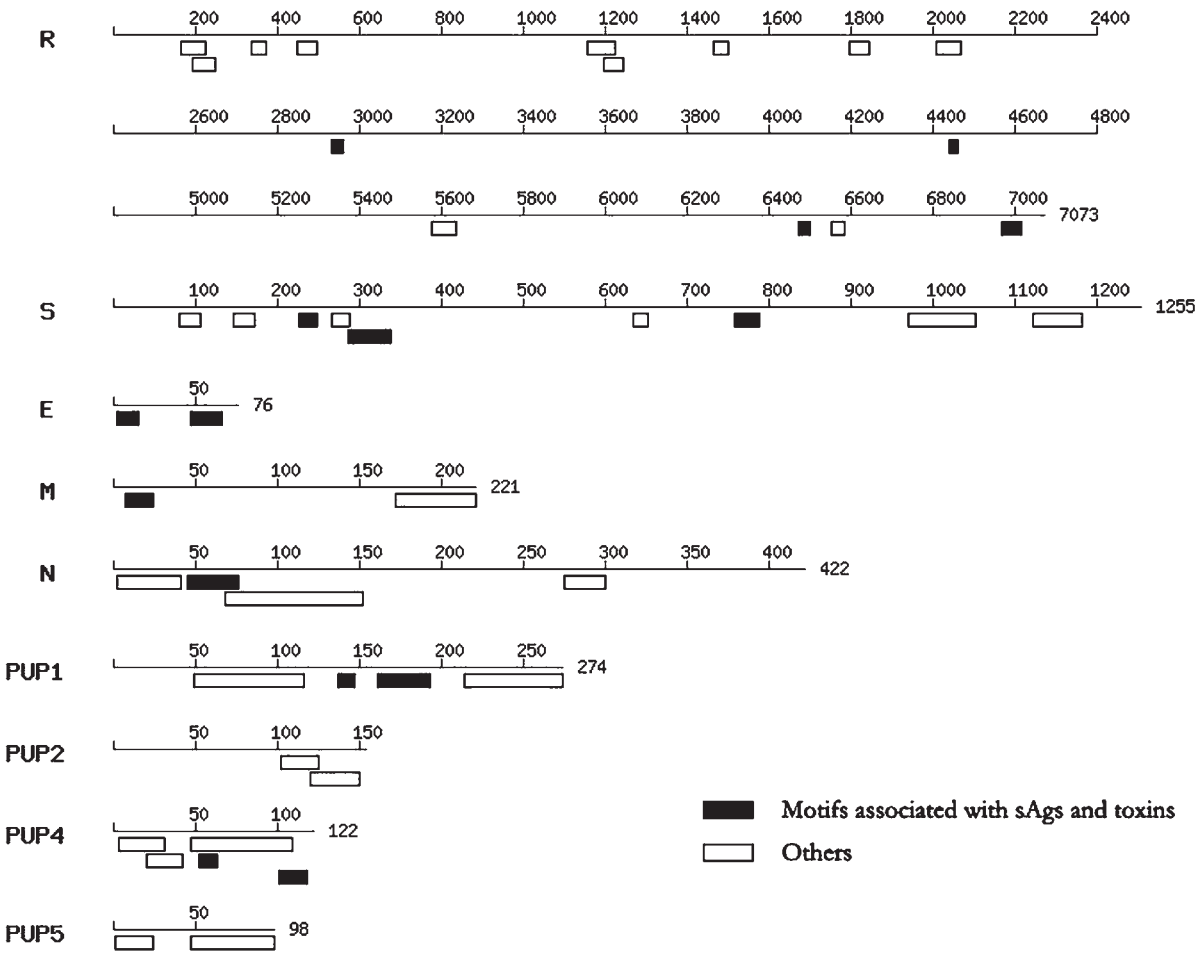

FIG. 1. The distribution of the predicted motifs in the SARS-CoV proteins compared to the local database.

Among these motifs, four were associated with neurotoxins and bullous pemphigoid antigen, overlapping with the predicted glycosylation sites at codons 227, 330, 783, 1116, and 1140. In the M protein, one motif possibly involved in sAg/toxin was located in the N-terminal exterior region, and another enzyme-like motif was located in the C-terminal interior region. Two motifs in the E protein were all involved in sAg/toxin. Two of the four motifs in the N protein were mutually clustered in the N-terminal region of 111 residues (45–152 a.a.), where a nuclear antigen sequence was identified. The fourth antigen-associated motif was located between residues 275-300 in the N protein.

The identified motifs in the four PUPs had dense distributions (Fig. 1). We noticed that the motifs at different places in the same proteins appeared to be related to the same toxic molecule, such as those in the PUP1 (Table 2). In other words, it is possible that several motifs in a SARS-CoV protein might cooperate to implement one function.

Classification of the motifs. 1. Motifs associated with sAg and/or toxin. These motifs in the SARS-CoV proteins can be divided into two groups. One group was similar to those from heat-labile enterotoxin and staphylococcal enterotoxin, exfoliative toxin, botulinum neurotoxin and bungarotoxin (a type of presynaptic neurotoxin), and cytotoxin 3 precursor; the other group was associated with proteases such as herpesvirus protease, hydrolase, and some proteases in *Escherichia coli* O157:H7 strain (Table 2). Fourteen motifs relating to sAgs and toxic molecules were identified (Fig. 1 and Table 2). Three motifs localized in the S protein were associated with neurotoxin, and two motifs identified in the E protein were both associated with botulinum neurotoxin, of which the N-terminal exterior one was similar to the type D precursor, and the C-terminal interior one was similar to the type B precursor. It suggested the possible functional relationship of them. One enterotoxin-like motif resided near an antigen site at the N-terminus of N protein. Moreover, a region of 85 amino acids was identified as an unclear antigen. Their antigenicities have been validated in recent study (8). In the middle of PUP1, two toxin-associated motifs were detected. Almost three-

TABLE 2. THE MOTIFS PREDICTED IN SARS-COV (BJ01 ISOLATE)

Protein	*Length (a.a.)*	*Position (a.a.)*		*Identity (%)*	*Motif length (a.a.)*	*Matched sequence*
		165	**227**	36	63	Putative enzymes [Escherichia coli O157:H7]
		193	**249**	34	57	Bacterial surface antigen family protein [Pseudomonas putida KT2440]
		336	373	36	38	Methionyl-tRNA synthetase
		448	498	33	51	Plaminogen activator, urokinase [Homo sapiens]
		1157	**1226**	31	70	P93 antigen
		1196	**1244**	38	49	Apoptotic protease activating factor-1 long isoform APAF-1L
		1465	1501	43	37	Anti-myosin immunoglobulin heavy chain variable region [Mus musculus]
R	7073	1797	1844	35	48	Interleukin-4 precursor (IL-4)
		2010	2071	34	62	Erythrocyte membrane-associated giant protein antigen
		2933	2960	39	28	Botulinum neurotoxin type F precursor (bont/F) (Bontoxilysin F)
		4441	4461	52	21	Heat-labile enterotoxin
		5577	5636	31	60	Putative glycosyl transferase [Escherichia coli]
		6475	6503	41	29	Mitogenic exotoxin Z-9 [Streptococcus pyogenes]
		6554	6584	48	31	Superoxide dismutase-blue shark
		6968	7019	31	52	Superantigen ypmc [Yersinia pseudotuberculosis]
		80	107	43	28	Coagulation factor II receptor; Thrombin receptor
		147	173	37	27	Intercellular adhesion molecule 1 precursor; CD54 [Homo sapiens]
		227	249	35	23	Hydrolase, presynaptic neurotoxin molecule: beta2-bungarotoxin
		266	**288**	44	23	Intercellular adhesion molecule 1 precursor; CD54 [Homo sapiens]
S	1255	**286**	**338**	32	53	Botulinum neurotoxin type G precursor (bont/G) (Bontoxilysin G)
		634	653	45	20	Epidermal growth factor [Mus musculus]
		759	789	39	31	Botulinum neurotoxin type G precursor (bont/G) (Bontoxilysin G)
		970	1052	31	83	Peptodoglycan recognition protein-like [Mus musculus]
		1123	1183	34	61	Bullous pemphigoid antigen 1-e [Mus musculus]
E	76	2	15	43	14	Botulinum neurotoxin type D precursor (Bontoxilysin D)
		47	66	35	20	Botulinum neurotoxin type B precursor (Bontoxilysin B)
M	221	7	24	39	18	Exfoliative toxin a
		172	221	37	50	Nitrate-inducible formate dehydrogenase-N alpha subunit [O157:H7]
N	422	2	41	30	40	SLP-76 associated protein ***Validated***
		45	**76**	38	32	Staphylococcal enterotoxin a
		68	**152**	31	85	Nuclear antigen 2 ***Validated***
		275	300	39	26	Antigenic virion protein [Human herpesvirus 6B]
		49	116	34	68	O-antigen polymerase [Shigella boydii]
PUP1	274	137	147	55	11	Botulinum neurotoxin type G precursor (Bontoxilysin G)
		161	193	34	33	Botulinum neurotoxin type C1 precursor (Bontoxilysin C1)
		214	274	33	61	Transcriptional regulator ume6
PUP2	154	**102**	**125**	38	24	Interleukin-2
		120	**150**	31	31	Macrophage inflammatory protein 3 alpha
		3	**31**	47	29	Similar to CD83 antigen [Mus musculus] [Rattus norvegicus]
		20	**42**	48	23	Tumor necrosis factor alpha; TNF alpha [Mus musculus]
PUP4	122	***47***	***109***	32	63	Lymph node homing receptor (Leukocyte-endothelial cell adhesion molecule 1)
		52	***63***	50	12	Botulinum neurotoxin type F precursor (Bontoxilysin F)
		101	***118***	44	18	Cytotoxin 3 precursor (Cardiotoxin analog III)
PUP5	98	1	24	38	24	Tumor necrosis factor (ligand) superfamily, member 7 [Mus musculus]
		47	98	31	52	Putative GDP-mannose-4,6-dehydratase; Ipsa [Caulobacter crescentus]

The overlap regions are indicated by bold font.

Table 3. Hits of the Selected Motifs in Six Viruses

	SARS-CoV	*OC43*	*PEDV*	*HRSV*	*HHerpV*	*HAdenoV*
Neurotoxin	8	17	2	20	27	29
Enterotoxin	4	2	18	13	10	13
IL-1	0	0	0	3	6	1
IL-2	17	0	0	0	3	2
Ig	12	7	1	273	16	18
TNF	16	0	0	0	35	20

Including the total hits matching to the local database.

fourths of the regions in PUP4 were covered by two successive toxin-like motifs near the N-terminal, overlapping a consensus sequence of lymph node homing receptor (Lnhr).

2. Motifs associated with cytokines. Eight types of cytokine-like motifs were detected (Table 2), which were associated with IL-2, IL-4, macrophage inflammatory factors, plasminogen activator and apoptotic protease activating factor, etc. Two thirds of PUP2 were covered by two motifs from the members of cytokines, and about a 30% region near the N-terminus of PUP5 was covered by a TNF-like motif. Identically, one TNF-like motif was also found in PUP4. Moreover, the existence of two motifs associated with transcription factors and polymerase suggested that PUP1 might have a relationship with transcriptional regulation.

3. Motifs associated with membrane surface molecules. Approximately five motifs were involved in the consensus segments of the surface antigen and receptor of the membrane molecules, which were mainly located in the structural proteins (Table 2). It should be noted that two motifs had a high similarity with intercellular adhesion molecule 1 (ICAM-1 near the N-terminal of the S protein, CD54), which is important in mediating immune and inflammatory responses (2). An Lnhr-like motif accounted for the major part of PUP4. The N protein, which is the component of nucleocapsid, contained a nuclear antigen-like motif. In the R protein, one region was found similar to that of a bacterial surface antigen family protein.

Comparison of the motifs between SARS-CoV and five other viruses. We have made a comparative analysis of five common viruses. Table 3 shows that each type of virus has a different composition of the observed motifs. Among the three kinds of coronaviruses, the SARS-CoV had more confident hits of IL-2 and TNF-motifs than the other two. On one hand, SARS-CoV and OC43 had more neurotoxin-like motifs than enterotoxin-like ones, on the other hand, PEDV showed high hits of enterotoxin-like motifs. The HRSV was characterized by high hits of Ig-like motifs. In summary, both IL-2 and cytotoxin-like motifs appeared in the SARS-CoV. This was different from OC43, PEDV and HRSV, but somewhat similar to HHerpV and HadenoV.

DISCUSSION

In this study, we identified 45 motifs that are similar to those known proteins relating to sAgs, toxins, cytokines and antigens, which have already been confirmed in previous studies. The sequences are relatively conserved across species, which suggest that similar functions may be played by these motifs (1). At the same time, the results strongly suggest that the distribution of the motifs is coincided with the function of the proteins. For instance, enzyme-like motifs were located in the R protein, adhere molecule-like motifs in the S protein, nuclear antigen-like motif in the N protein, and the motifs associated with transcriptional regulator in the PUP1. Thus, identification of these motifs provides active evidence to analyze and understand the functions of relevant regions in the proteins of SARS-CoV.

The whole set of motifs in SARS-CoV is different from those of OC43, PEDV, HRSV, HHerpV and HAdenoV. Each set of motifs in different viruses can determine different immune reactions.

The motifs of antigenicity and toxin play a key role in pathogenesis of SARS. Similar to many other RNA viruses, the SARS-CoV has some common features, but there are some other features of infecting human body and causing disease, determined by SARS-CoV's special viral proteins. The motifs, as small conserved region within a large biological sequence, are essential to the function of the viruses. Both structural and non-structural proteins of the SARS-CoV practically have biological activities. It is not difficult to understand what would happen while many toxic and antigenic motifs of the active proteins are exposed to the host body.

The data show that the SARS-CoV possesses many motifs associated with sAgs, toxins and cytokines. The toxic motifs are mainly involved in neurotoxin, entero-

toxin, cytotoxin and some proteases. The cytokine-like motifs are related to inflammatory factors, apoptosis factors and TNF.

The sAgs are powerful microbial toxins that target the host immune system by directly binding the MHC-II and T cell receptor (TCR), without requiring the APC processing (4). Since Hillyard first discovered that a group of viral sAg (minor lymphocyte stimulating antigens) could generate strong T-cell proliferative response, more and more evidence suggest that the viral sAgs are involved in immune-mediated diseases (14). Unlike normal peptide antigens that only stimulate between 0.001% and 0.0001% of T cells, many types of sAgs could activate up to 20% of all T cells. Even though coming from the same ancestral gene, the same type of sAgs may have variant structures, for instance, the staphylococcal and streptococcal enterotoxins, despite they keep the similar virulence (4,9).

The existence of sAg-like motifs in the SARS-CoV provides clear evidence to interpret how the virus causes the disease. The sAg and toxic molecules on the PUPs and S, E, N proteins could bypass the normal immune passway and directly stimulate the T cell. In the case that no costimulating signal takes part in the procedure, the T cell could not be fully activated and trend to apoptosis, and it would cause markedly decrease of T-cell level in the blood, especially $CD4^+$ and $CD8^+$ cells. Consequently, it makes the massive inflammatory factors release in a short time and induce the autoimmunity (5,10). Furthermore, the danger is that the human alveolar cells are widely attacked by self-cytokines besides probable damage, partly from the virion, followed by fever, muscle pain, pulmonary edema and allergic angitis. The enterotoxin motifs on the S protein and PUP synthesized by SARS-CoV can sometimes lead to diarrhea, but it seems to be less serious than PEDV-caused illness in swine, which have more enterotoxin motifs (Table 3).

Unlike HRSV, OC43 and PEDV, the special IL-2 and TNF-like motifs of the SARS-CoV must play an exclusive role in virulence. It is reported that the TNF-a as inflammatory cytokines is associated with bronchial hyper responsiveness by reducing the response to ·-agonists and increasing the reactivity to methacholine, the airway neutrophils and alveolar macrophages (9). These motifs located in PUP4 and PUP5 might act as the allergens.

We also focused on PUP4, which contains a toxic motif similar to TNF, and other ones similar to endothelial cell adhesion molecule. It is postulated that PUP4 might positively take part in the interference with the host's normal immune reaction and result in the immune disorder.

The N protein, as the component of nucleocapsid, has an outstanding expression of antigen-associated motifs, which implies the important essential of its antigenicity.

The ICAM-1-like motif in the S protein has a crucial role in targeting and invading the host cells with special affinity, which is supposed to mediate the virion adhering to the alveolar epithelial cells (11), intestines epithelial cells and brain cells etc.

The virion compactly integrates multiple motifs may play complex roles in making use of the host immune system and interfering in the normal immune regulation to counterattack the host. This is not dramatic but factual, for the construction of viral proteins is so perfect and ingenious.

The signification in the vaccine and drug design, diagnosis, and therapy. In summary, the recognition of the SARS-CoV's structure not only help us understand the molecular basis of its virulence and pathogenesis, but also provide us detailed information and new approaches to develop more specific diagnostics and effective therapeutics against SARS. Based on this study, it is essential to avoid the toxin structure and select the special antigenic determinant during the vaccine preparation. The accuracy of the molecular probe should be modified in order to improve the diagnostic preciseness. When we check the cytokines in the blood, the potential errors caused by the cross-reaction between SARS motifs and antibodies should not be underestimated. One of the important therapeutic strategies is to block the viral track against the host immune system, so the anti-toxin antibodies should be a good selection to treat this disease.

ACKNOWLEDGMENTS

We thank Ministry of Science and Technology of China, Chinese Academy of Sciences, and National Natural Science Foundation of China for financial support. We are also grateful to collaborators and clinicians from Peking Union Medical College Hospital, National Center of Disease Control of China, and the Municipal Governments of Beijing, Chongqing, and Hangzhou.

REFERENCES

1. Alcami, A. 2003. Viral mimicry of cytokines, chemokines and their receptors. Nat. Rev. Immunol. 3:36–50.
2. Anderson, M.E., and T.J. Siahaan. 2003. Targeting ICAM-1/LFA-1 interaction for controlling autoimmune diseases: designing peptide and small molecule inhibitors. Peptides 24:487–501.
3. Drosten, C., S. Gunther, W. Preiser, et al. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 348:1967–76.
4. Fraser, J., V. Arcus, P. Kong, et al. 2000. Superantigens - powerful modifiers of the immune system. Mol. Med. Today 6:125–32.

5. Friedman, S.M., J.R. Tumang, and M.K. Crow. 1993. Microbial superantigens as etiopathogenic agents in autoimmunity. Rheum. Dis. Clin. North Am. 19:207–22.

6. Ksiazek, T.G., D. Erdman, C.S. Goldsmith, et al. 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348:1953–66.

7. Li, T., Z. Qiu, Y. Han, et al. 2003. Rapid loss of both $CD4^+$ and $CD8^+$ T lymphocyte subsets during the acute phase of severe acute respiratory syndrome. Chin. Med. J. (Engl.) 116:985–7.

8. Lin, Y., X. Shen, R.F. Yang, et al. 2003. Identification of an epitope of SARS-coronavirus nucleocapsid protein. Cell. Res. 13:141–5.

9. Llewelyn, M., and J. Cohen. 2002. Superantigens: microbial agents that corrupt immunity. Lancet Infect. Dis. 2:156-62.

10. McCormick, J.K., J.M. Yarwood, and P.M. Schlievert. 2001. Toxic shock syndrome and bacterial superantigens: an update. Annu. Rev. Microbiol. 55:77–104.

11. Popper, H.H., S. Pailer, G. Wurzinger, et al. 2002. Expression of adhesion molecules in allergic lung diseases. Virch. Arch. 440:172–80.

12. Poutanen, S.M., D.E. Low, B. Henry, et al. 2003. Identification of severe acute respiratory syndrome in Canada. N. Engl. J. Med. 348:1995–2005.

13. Qin, E., J. Wang, W. Li, et al. 2003. A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01). Chinese Sci. Bull. 48:941–948.

14. Torres, B.A., and H.M. Johnson. 1998. Modulation of disease by superantigens. Curr. Opin. Immunol. 10:465–70.

15. Tsang, K.W., P.L. Ho, G.C. Ooi, et al. 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. N. Engl. J. Med. 348:1977–85.

16. Ziebuhr, J., E.J. Snijder, and A.E. Gorbalenya. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. 81:853–79.

Address reprint requests to:
Dr. Huanming Yang
Beijing Genomics Institute (BGI)
Chinese Academy of Sciences (CAS)
Beijing 101300, China

E-mail: yanghm@genomics.org.cn

Received December 12, 2003, accepted March 19, 2004.