The M Protein of SARS-CoV: Basic Structural and Immunological Properties

Yongwu Hu[1,2]*, Jie Wen[1]*, Lin Tang[1]*, Haijun Zhang[1]*, Xiaowei Zhang[1], Yan Li[1], Jing Wang[1,3], Yujun Han[1], Guoqing Li[1], Jianping Shi[1], Xiangjun Tian[1,2], Feng Jiang[1], Xiaoqian Zhao[1], Jun Wang[1], Siqi Liu[1], Changqing Zeng[1], Jian Wang[1,2], and Huanming Yang[1,2]#

[1] Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China; [2] James D. Watson Institute of Genome Sciences, Zhijiang Campus, Zhejiang University and Hangzhou Genomics Institute, Hangzhou 310008, China; [3] College of Life Sciences, Peking University, Beijing 100871, China.

We studied structural and immunological properties of the SARS-CoV M (membrane) protein, based on comparative analyses of sequence features, phylogenetic investigation, and experimental results. The M protein is predicted to contain a triple-spanning transmembrane (TM) region, a single N-glycosylation site near its N-terminus that is in the exterior of the virion, and a long C-terminal region in the interior. The M protein harbors a higher substitution rate (0.6% correlated to its size) among viral open reading frames (ORFs) from published data. The four substitutions detected in the M protein, which cause non-synonymous changes, can be classified into three types. One of them results in changes of pI (isoelectric point) and charge, affecting antigenicity. The second changes hydrophobicity of the TM region, and the third one relates to hydrophilicity of the interior structure. Phylogenetic tree building based on the variations of the M protein appears to support the non-human origin of SARS-CoV. To investigate its immunogenicity, we synthesized eight oligopeptides covering 69.2% of the entire ORF and screened them by using ELISA (enzyme-linked immunosorbent assay) with sera from SARS patients. The results confirmed our predictions on antigenic sites.

Key words: SARS-CoV, the M protein, enzyme immunoassay, antigenicity

Introduction

In the extensive processes of evolution, many viruses acquire a lipid bilayer from their host cells. A special protective and functional mosaic structure is thereby formed with the viral structural proteins that are integrated into the membranous envelope. The viral membrane proteins often resemble their host counterparts in the structure and functions (*1*).

The M (membrane) protein (also known as E1 membrane glycoprotein or matrix protein) is one of major membrane proteins of the coronavirus together with the S (spike) and the E (envelope) proteins (*2*). It has been postulated that the M protein is related to viral infectivity through binding to the viral S protein and the host surface receptor(s), thus promoting membrane fusion (*1*, *3*). It appears to be involved in the antigenicity demonstrated by the virus-induced immune responses of the host. As the most abundant protein in the virion of coronaviruses, the M protein may also be one of the key components in viral assembly and morphogenesis, involved in regulation of replication and packing the genomic RNA into viral particles (*4*, *5*).

We report herein the comprehensive analyses of the M protein based on the sixteen first reported SARS-CoV genome sequences, as well as seventeen members of *Coronaviridae* and 361 other single-strand RNA (ssRNA) viruses.

*** These authors contributed equally to this work.**

 # Corresponding author.

E-mail: yanghm@genomics.org.cn

Results

Genomic features of the M protein

The ORF of the M protein, which has a defined 5′ transcription regulation sequence (TRS) (5′-CUAAAC-3′), is located between the ORFs for the E protein and PUP3 (putative uncharacterized protein 3) (*6*). The ORF is 666 nucleotides (nt) long, encoding a protein of 220 (+1) amino acid (a.a.) residues with a predicted molecular weight of 25.06 KD. It has an overall GC content of 44.5% (A: U: C: G = 164: 200: 148: 153 = 24.66%: 30.07%: 22.25%: 23.00%), significantly higher than that of the genome average (40.8%) (Figure 1-I). Its higher GC content is largely due to the higher fraction of hydrophobic amino acids. The amino acid composition of the M protein is summarized in Table 1.

One of the noticeable features of the M protein is that it has a relatively low fraction of Cys (3/221, 1.4%) in comparison with other defined viral proteins, except the N (nucleocapsid) protein that is essentially cysteine-free. Four residues, namely Leu, Ala, Ile, and Val, contribute to about 40% of the protein, each occupying 14.0%, 8.6%, 8.1%, and 7.2%, respectively. The overall M protein is theoretically basic (pI 9.63) and positively charged (Arg 6.8% and Lys 2.7%; Table 1).

Table 1 The Biochemical Features of the M Protein

Feature	No. of a.a.		Percent (%)
Non-polar, neutral	Leu	31	14.00
	Ala	19	8.60
	Ile	18	8.10
	Val	16	7.20
	Gly	15	6.80
	Phe	11	5.00
	Trp	7	3.20
	Met	7	3.20
	Pro	5	2.30
Neutral polar	Thr	13	5.90
	Asn	13	5.90
	Ser	12	5.40
	Tyr	9	4.10
	Gln	5	2.30
	Cys	3	1.40
Positive polar	Arg	15	6.80
	Lys	6	2.70
	His	3	1.40
Negative polar	Asp	6	2.70
	Glu	7	3.20
Total		221	100
pI value	9.3		
Hydrophobicity (%)	40.70		
Hydrophilicity (%)	36.20		
Charge (+)(%)	10.90		
Charge (-)(%)	5.90		

Subregional structure of the M protein

The M protein is characterized by a typical TM (transmembrane) region composed of three putative TM domains of 80 a.a. residues that account for about one third of the entire protein. It is located between Codons 19-98 in the ORF. Its overall features are summarized in Table 1 and Figures 1 and 2. The first TM domain (TMI) is composed of 19 a.a. residues (Codons 19-37). The non-polar, neutral amino acids account for a substantial fraction (83%) of the total domain, making it strongly hydrophobic (Figures 1 and 2). The second domain (TMII) is between Codons 50-72 (23 residues), with 81% non-polar, neutral residues, and with physiochemical features similar to TMI. A lipoprotein attachment site for prokaryotic membrane is identified at Codon 53 in TMII. The first inter-TM

segment between TMI and TMII has a high probability of being located in the interior. A signal peptide of 39 a.a. residues in length with a likely cleavage site between Codons 39 and 40 (AYS-NR) is predicted within the first inter-TM segment between TMI and TMII. The TMIII domain is located between Codons 76-98 (23 residues), with 68.2% non-polar, neutral residues, and thus is not as hydrophobic as the other two. The segment between TMII and TMIII is possibly located in the exterior. A lipoprotein attachment site for prokaryotic membrane is predicted at Codon 75 in this segment or at its boundary with TMIII.

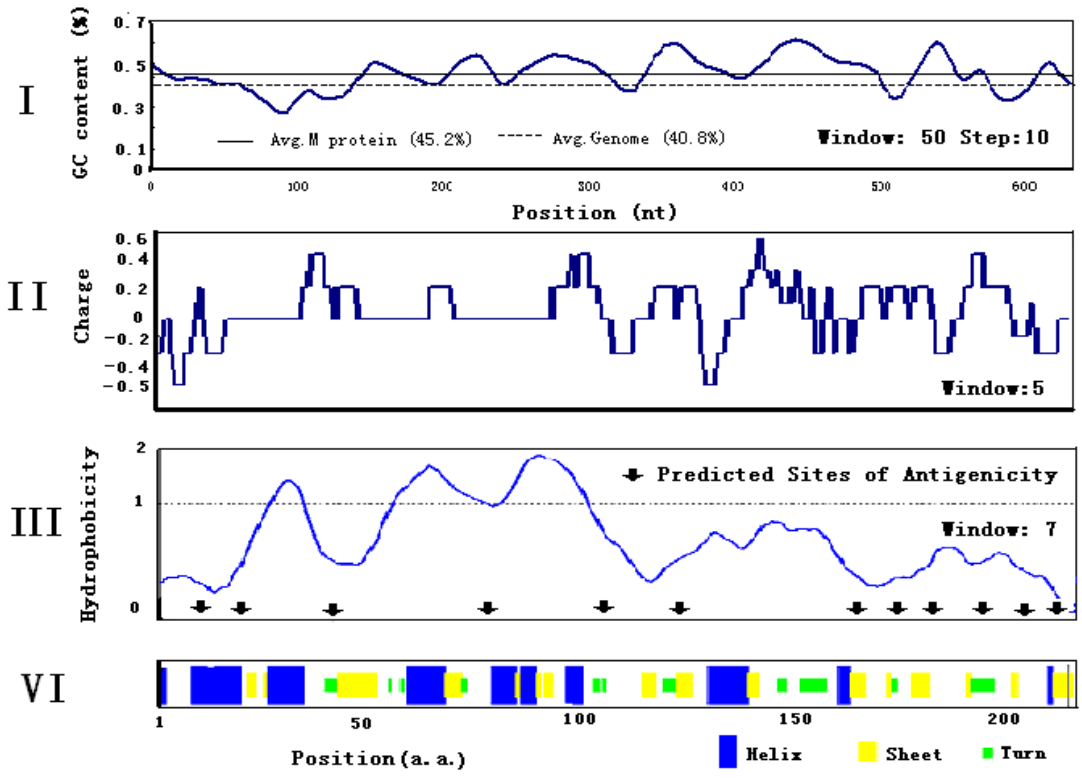

Fig.1 The distribution of GC content (I), charge (II), the predicted hydrophobicity and antigenicity (III) and secondary structures (IV) of the M protein.

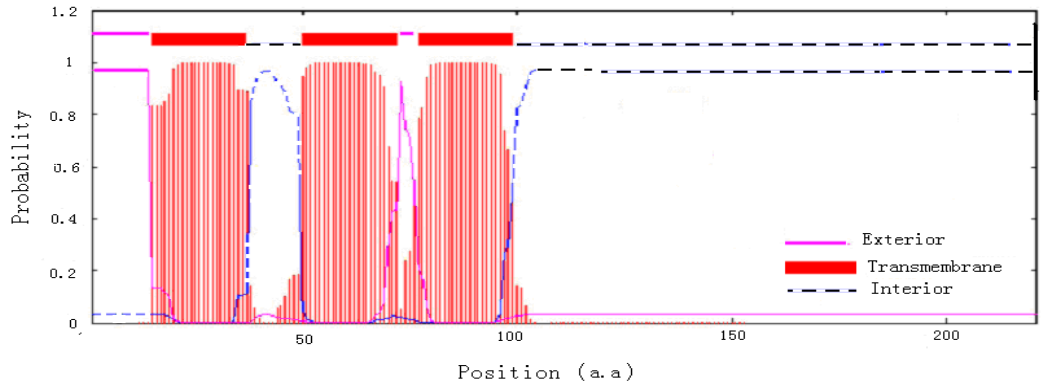

Fig.2 The predicted transmembrane domains of the M protein.

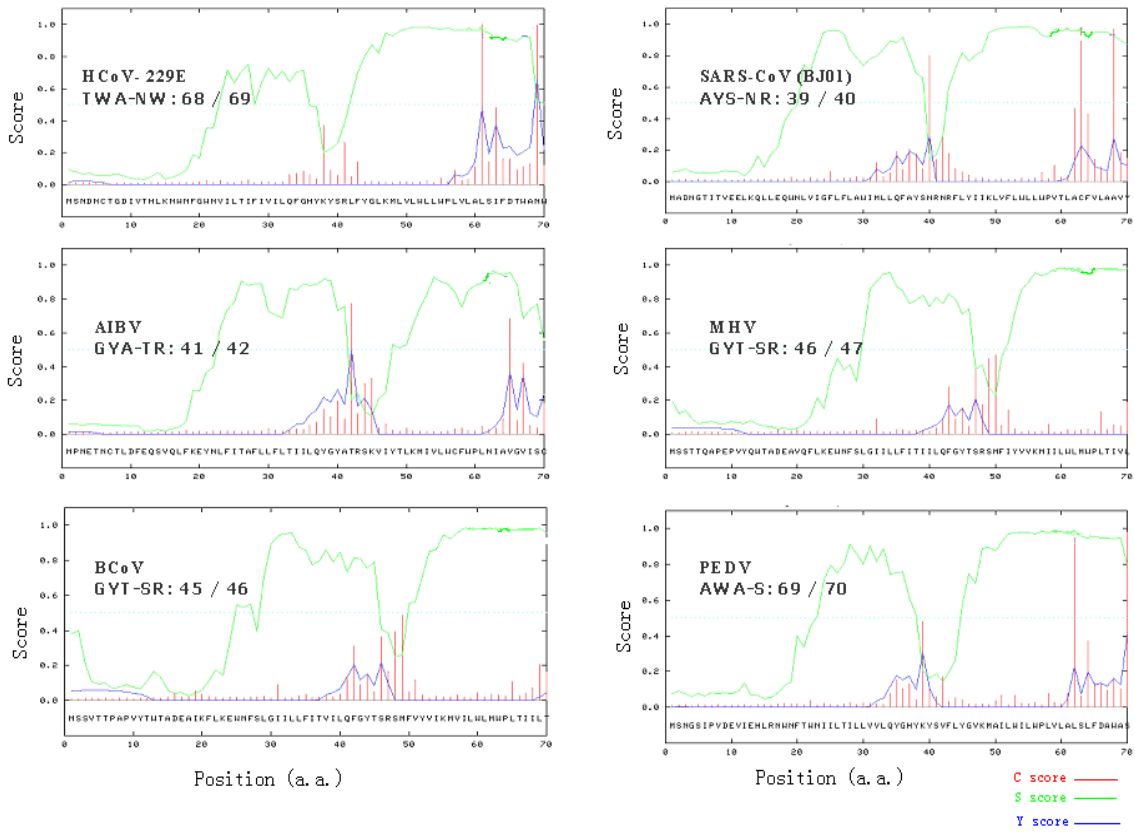

Fig. 3 Predicted signal peptides and cleavage sites of the M protein in six coronaviruses.

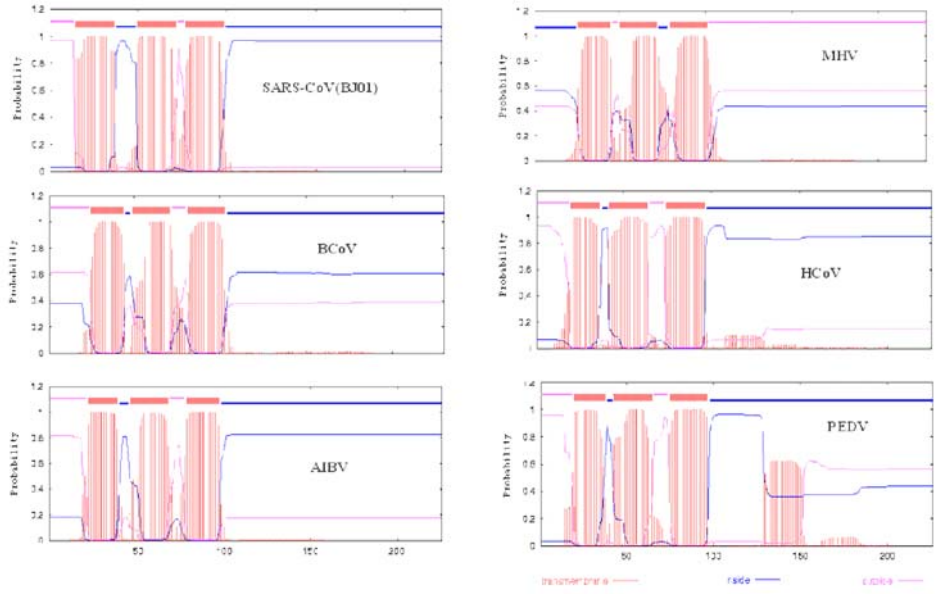

Fig. 4 The predicted transmembrane domains in six coronaviruses.

The amino-terminal exterior region is composed of only 18 codons (including start codon), and is locally acidic (pI 4.20; hydrophobicity 33.3%; hydrophilicity 55.6%). A single putative glycosylation site is found at Codon 4 (NGTI). The carboxyl-terminal region is composed of 123 residues (56.1%) and is located in the interior. Three protein kinase C phosphorylation sites (SFR at Codon 98, TSR at Codon 171, SQR at Codon 183) and three N-myristoylation sites (GTIVTR at Codon 125, GTDSGF at Codon 187, GSNDNI at Codon 211) were predicted in this region. The putative glycosylation sites in the interior, as appearing in the M protein of HcoV-229E and PEDV, might be considered in relation to the orientation, even though it is very unlikely active (Figures 2, 4 and 5).

Variations of the SARS-CoV M protein and its orthologues in Coronaviridae

No significant homology has been detected by comparative analysis of the published genome sequences from GenBank, but high similarity is identified for the M protein orthologues in all known members of *Coronaviridae* (Figure 5). An unrooted phylogenetic tree is proposed on the basis of the M protein, which places the SARS-CoV between Groups I and III, but still distantly related to each other (Figure 6).

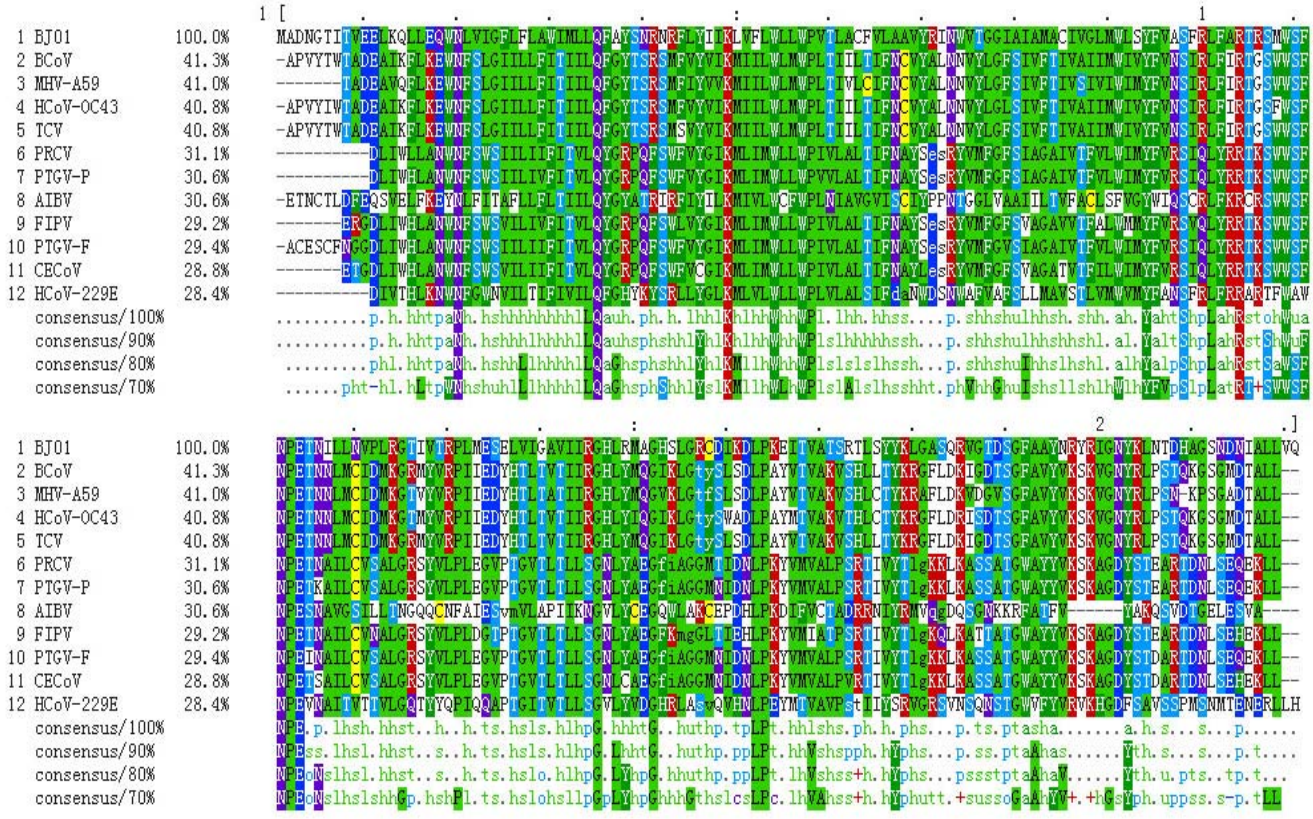

Fig. 5 The multiple alignments of amino acid sequences of the M protein in 12 coronaviruses.

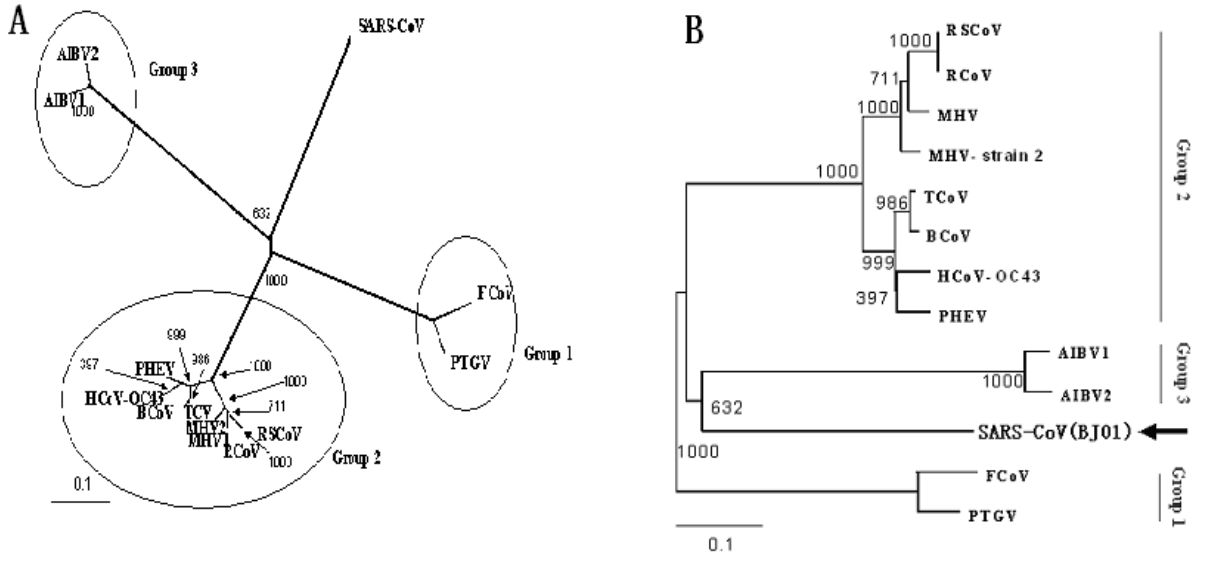

Fig. 6 A phylogenetic tree of the coronaviruses based on the M protein. A: ClustalW, output by Treeview. B: ClustalW, output by Phylodraw. AIBV: avian infectious bronchitis virus; BCoV: bovine coronavirus; CCoV: canine coronavirus; FCoV: feline coronavirus; FIPV: feline infectious peritonitis virus; HCoV-229E: human coronavirus 229E; HCoV-OC43: human coronavirus OC43; MHV: murine hepatitis virus; PEDV: porcine epidemic diarrhea virus; PHEV: porcine hemagglutinating encephalomyelitis virus; PRCoV(PRV): porcine respiratory coronavirus; PTGV: porcine transmissible gastroenteritis virus; RCoV: rat coronavirus; RSCoV: rat sialodacryoadenitis coronavirus; SARS-CoV: the human SARS-associated coronavirus; TCoV (TVC): turkey coronavirus.

Four substitutions in the M protein have been detected from sixteen isolates of the published SARS-CoV genomes (Table 1). The substitutions are distributed through all the 3 predicted regions: a G/A transition (Glu/Lys) at Codon 11 (SIN2500) in the exterior, a T/G transversion (Phe/Cys) at Codon 27 (CUHK_Su10) in TMI, a C/T transition at Codon 68 (HKU_39849) in TMII, and a T/C transition at Codon 154 (URBANI) in the interior, respectively. All substitutions are predicted to lead to amino acid alterations and changes of physical and/or chemical features (Table 2).

Table 2 Predicted Local Changes of Biochemical Features by Substitution

Isolate	Location	Position a. a. (nt)	Substitution a. a. (nt)	pI value	Hydrophobicity (%)	Hydrophilicity (%)	Charge (+) (%)	Charge (-) (%)
SIN2500	Exterior (1-18)	11 (26,409)	Glu/Lys (G/A)	4.20-4.80	(33.3)	(55.6)	5.6-11.1	22.2-16.7
CUHK_Su10	TM I (19-37)	27 (26,458)	Phe/Cys (T/G)	(5.60)	73.7-68.4	(10.5)	0	0
HKU_39849	TM II (50-72)	68 (26,581)	Ala/Val (C/T)	(7.90)	63.6-68.2	(9.1)	(4.5)	0
URBANI	Interior (99-221)	154 (26,838)	Ser/Pro (T/C)	(9.80)	(30.9)	45.5-44.7	(15.4)	(7.3)

ELISA tests of the M protein

To investigate the antigenicity properties of the M protein, eight synthetic oligopeptides were tested by the enzyme-linked immunosorbent assay (ELISA). The detailed absorbance data determined by ELISA for these oligopeptides are illustrated in Table 3. In this experiment, sera collected from 11 SARS patients from Guangdong Province and Beijing in Mainland China were used to measure the immune reactivity. Normal sera were used as negative controls and SARS-CoV lysatea as positive controls. The ELISA test result in a 96-well plate is showed in Figure 8.

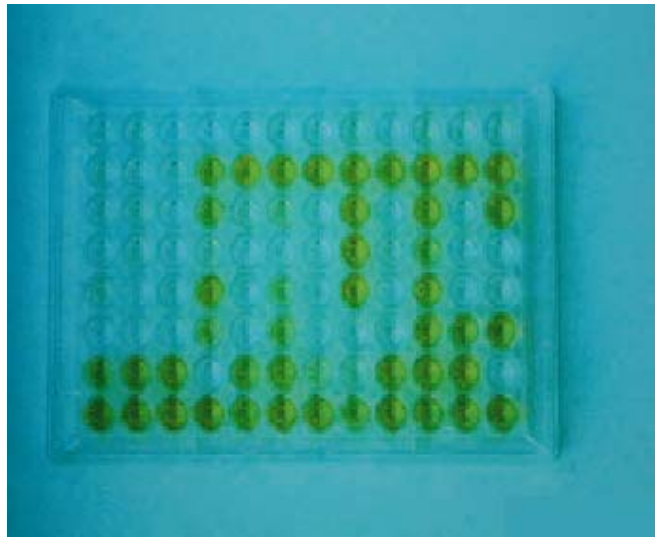

Fig. 7 The immunoassay of oligopeptides in a 96-well plate. The sample distribution listed in Table 3.

Table 3 The ELISA result of synthetic oligopeptides

Sample name	GD01	GZ04	GZ10	ZX10	ZX03	ZX26	ZX20	BJ01	BJ02	BJ03	BJ04	Negative control
M-001	0.498	0.167	0.668	0.505	0.040	0.032	0.053	1.394	0.235	0.149	0.453	0.014
M-017	0.041	0.023	0.024	0.104	0.022	0.019	0.013	0.092	0.047	0.062	0.077	0.019
M-103	0.076	0.032	0.038	0.081	0.017	0.016	0.013	0.068	0.063	0.073	0.088	0.035
M-119	0.288	0.148	0.037	0.027	0.075	0.047	0.049	0.083	0.030	0.097	0.088	0.013
M-137	0.706	0.668	0.857	0.707	0.117	0.126	0.103	0.882	0.580	0.780	0.880	0.111
M-162	0.139	0.076	0.049	0.146	0.018	0.028	0.019	0.241	0.180	0.153	0.241	0.050
M-189	0.128	0.085	0.331	0.030	0.138	0.126	1.622	0.259	0.837	0.294	0.515	0.021
M-206	0.129	0.102	0.168	0.171	0.151	0.743	0.055	0.764	0.670	0.335	0.579	0.020

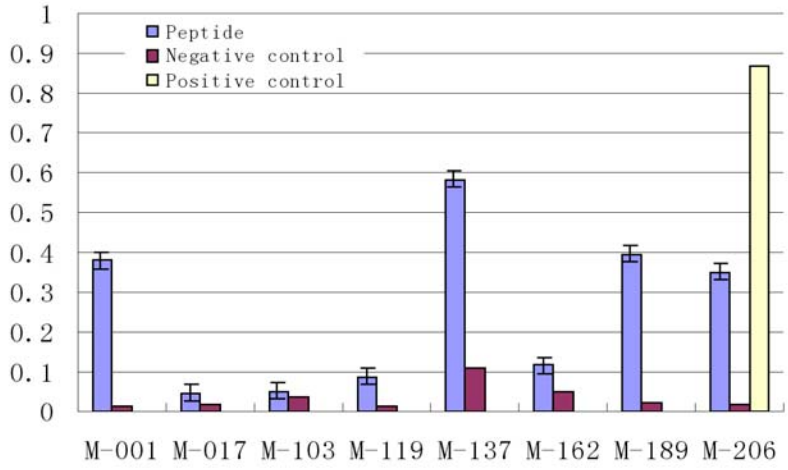

Fig. 8 The average immune reactivity of 8 oligopeptides by ELISA. Four oligopeptides show higher absorbance, suggesting immune reactivity with sera from SARS patients. Synthetic oligopeptides cover all the three predicted subregions of the M protein except for partial TM regions with extremely high hydrophobicity and low predicted antigenicity.

Discussion

The M protein is the core membrane protein in SARS-CoV

Previously published experimental results and bioinformatics predictions have provided strong evidence that the M protein found in other coronaviruses is a structural membrane protein (*4*). The M protein of SARS-CoV demonstrates relatively high homology at the amino acid sequence level with its orthologues in *Coronaviridae*. The high similarity of the M proteins in primary (that is, amino acid sequence) and secondary structure suggests the phylogenetic relationship between SARS-CoV and other members of *Coronaviridae*, conservativeness in evolution, and its importance to the survival of the virus.

Our results suggest the existence of three predicted domains in the M protein, and indicate that the orientation of the TM region has its N-terminus in the exterior of the virion and its C-terminus in the interior. However, the determination of its orientation is more difficult but crucial to the subsequent function analysis.

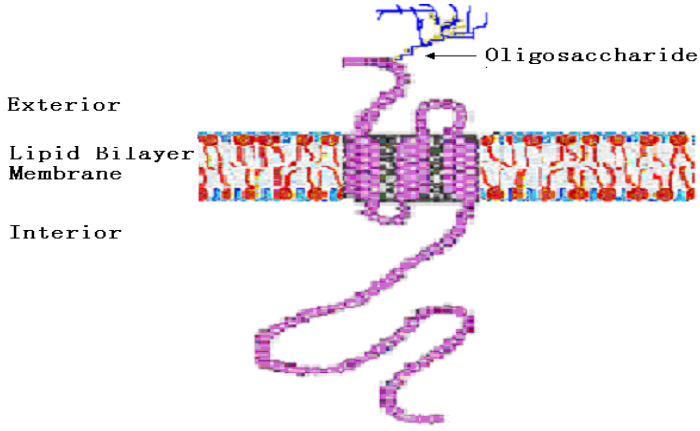

Fig.9 A diagram of the predicted TM structure in the SARS-CoV M protein.

An important discrepancy between biological and bioinformatic data was encountered during our analysis with the TM prediction software. Results of molecular biological and medical experiments strongly suggest that the M protein in MHV has its N-terminus in the exterior (*4*). But TMHMM, currently rated as the best software for TM prediction (*7*), yields a reverse orientation with the N-terminus in the interior. To rule out possible artifacts, we applied several currently available software packages to predict the TM orientation of the M protein. All other software closely replicated our results with TMHMM and predicted a similar structure of the corresponding regions in the M protein. The subregionally predicted orientation of the first and second inter-TM segments, as well as the predicted conformation of an α-helical structure and other features identified in all three TM domains, not only provide strong evidence for the existence of the TM structure, but also support the predicted orientation.

Several predicted features are related to the TM region of the M protein. The predicted signal peptide has a structure and position similar to the M protein in other coronaviruses, posing a question as to whether it is actually cleaved and released. The cleavage site (AYS-NR, between Codons 39 and 40) may be found within the first inter-TM segment (12 residues between Codons 37-50), which is very likely to be in the interior, thus reducing the possibility of cleavage by any cellular proteinase activities outside the virion. However, the possibility of the cleavage and release before viral assembly or after membrane fusion should be considered.

The carboxyl interior region is responsible for most functions and antigenicity of the M protein

The M protein appears to be involved in the regulation of replication and packing of genomic RNA into

viral particles (*4*, *8*). The M protein has the longest interior region (123 residues) inside the virion, much longer than that of the S (36 residues) and E (39 residues) proteins. It is reasonable to postulate that the M protein is the core membrane protein responsible for a major function inside the virion.

The C-terminal interior region is characterized by higher subregional GC content (47.6%) than the average, high pI (pI 9.80) and high hydrophilicity (45.5%), being overall positively charged with subregional variations, and large size in the virion. This region is predicted to have at least two phosphorylation sites for casein kinase II (TSR at Codon 171, SQR at 183), which might be related to its possible interaction with the interior of the S and E, as well as N proteins.

A segment (YFV-S- - L- -R-TSMWSFNPE), located between Codons 94 and 114, is identified as partially overlapping with the interior membrane boundary of TMIII (Codons 76-98), which is the most conserved (consensus 70-100%) motif in the M protein among all coronaviruses. It has been suggested that the region containing this consensus could be related to interaction with the N protein and viral RNA (*9*). We also have found another segment, PETNILLNVP, with two residues overlapping with the C-terminal part of the motif discussed above (—TSMWSFNPE). The blast results reveal that the amino acid sequence of this motif is identical (nt homology = 24/25, identity = 96%) to a segment in Vomeronasal 1 receptor (V1rf2) of mice (AY065513). V1rf2 is a well-known pheromone receptor on the olfactory bulb in mice. It functions in membrane trafficking involved in the transduction of cellular Ca^{2+} signals to the existing downstream effectors(10). This discovery indicates that the motif might play a crucial role in infecting epithelial cells in the respiratory tract, suggesting a possible link to a respiratory manifestation in SARS.

We have searched for the possible cleavage sites for the endogenous proteinase activity that would function inside the virion. No candidate site has been defined for the main viral proteinase $3CL^{pro}$ (chymotrypsin-like protease). However, four predicted sites (Ala|Gly, Gly|Gly, Gly|Aln) for the accessory proteinase PL^{pro} (papain-liker proteinase) have been identified at Codon 78 in TMIII and at Codons 151, 201 and 210 inside the C-terminus. This would suggest a possible autoproteolytic process in the M protein at certain stages. But, blast searches did not yield any functional information about the defined subregions of this protein.

At least eight predicted antigenic sites have been localized in this region. It has been postulated that these sites inside the virion might be related to the viral antigenicity after endocytosis when the virion is collapsed and the non-exterior regions of the M protein molecules are released into the host cytoplasm and exposed, or immediately after translation and subsequent modification before integrating into the host membranous components. The immunological results were generally consistent with the predicted antigenic regions and basic structure analysis of the M protein with bioinformatics tools. Four segments, 1-20, 137-157, 189-211, 206-221, were showed to react with antibodies of SARS patients. These segments, as predicted by bioinformatics tools, exhibit high hydrophilicity. The 1-20 segment is an exterior region, and the others are interior regions. The predicted TM region with high hydrophobicity was found to be non-immunoreactive experimentally, and our findings conformed well to results obtained by using bioinformatics tools.

Function-related structure of the N-terminal exterior region

Glycoproteins on the viral surface are known to be generally involved in antigenicity demonstrated by the viral pathogens and immunoreponse by the host. The M protein might be related to the viral infectivity through binding to the viral S protein, and then subsequently binding to the host cellular receptor(s) and helping with the membrane fusion (*1*, *3*, *11*). The N-terminal exterior region of the M protein is postulated to be responsible for its important roles in the pathogen-host interaction and membrane fusion with the host cell. The predicted presence of a single putative N-glycosylation site at the N-terminus (Codon 4) is highly conserved in other coronaviruses at similar positions and surroundings. The glycosylation site is classified as Type N-glycosylation (linked with Asn). As with other coronaviruses in Group I and II, no O-linked oligosaccharide site has been identified (*12*).

In an MHV experimental system, the monoclonal antibodies raised against the M protein were able to neutralize infectivity in vitro, thus protecting the host mice against lethal challenge (*3*, *11*). Two antigenic sites have been detected in the exterior region and

could be potentially involved in virus-host interaction supported by our experimental data. An unexpected finding was that all 10 isozymes of the caspase family, which, together with granzyme and thrombin, are well known to be involved in cell apoptosis, did not have any cleavage sites in the entire M protein. It should be noted that the E protein can induce apoptosis (*13*), and our data suggest that there may be some protective mechanism in the virus to prevent apoptosis.

Three different types of substitutions found in the M protein

One of the main discoveries of this study is that the M protein has the highest substitution rate among all the defined non-structural and structural proteins in SARS-CoV. Comparative analysis with genome sequences of all the 16 isolates has detected a total number of 4 nucleotide substitutions in the ORFs for the M protein, with an estimated overall substitution rate of 0.6% (4/666 nt). This rate is much higher than that of the ORFs for any other viral structure proteins, and even higher than the average rate of the entire genome (0.46%). It should be emphasized that all the mutations were only detected from one of the isolate genomes, and should not be taken as the equivalent of the evolution rate. We have re-estimated this rate with a simple alternative: since 16 sequences of the M protein have been used, a reasonable approximation for this mutation rate might be 0.03% (4/666×16). Using this approach, the average rate for the entire viral genome would be 0.02%, and that for the S protein would be 0.025%.

The four substitutions we have discovered are all non-synonymous, which suggests that SARS-CoV might gain advantages from possible positive selection related to its growth property, host range, or ability to escape from host immuno-surveillance. The four substitutions might represent three different classes of substitutions that lead to different changes in the biochemical features of the region, as described below.

1. Substitution leading to the change of pI and charge, and probably related to antigenicity.

The first substitution is identified in the amino-terminal exterior region, at nucleotide position 26,409 of the entire genome and Codon 11 of the ORF for the M protein. It is a transition (G/A), leading to a residue substitution from Glu to Lys. Both are strongly polar and absolutely hydrophilic (100%) amino acids. However, the former is well known to be highly acidic (pI 4.50) and negatively charged, while the latter highly basic (pI 8.50) and positively charged. Therefore, the positional residue remains absolutely hydrophilic (100%), but alters the pI and charge of the protein. Consequently, a significant subregional alteration of pI (pI 4.20- 4.80) and charge (+/- from 5.6/22.2 to 11.1/16.7) caused by the substitution could be predicted for the exterior region (Codons 1-18) of the virion. It is known that high pI value generates a stronger antigenicity, and thus enhanced antigenicity should be postulated for this substitution in the M protein.

2. Substitution leading to the change of hydrophobicity, and related to membrane stability.

Two other substitutions have been identified, each from a single isolate, in TMI and TMII, respectively. The former (T/G) is a transversion that leads to a residue substitution from Phe to Cys. The latter is a C/T transition leading to residue substitution from Ala to Val. Both Phe and Cys are neutral and weakly acidic with the same pI (pI 5.6). Phe is non-polar (zero charge) and absolutely hydrophobic, whereas Cys is polar and also absolutely hydrophobic. It is expected that the subregional pI (pI 5.60), non-polarity (zero charge) and hydrophilicity (10.5%) will remain unchanged, but the hydrophobicity of TMI might be affected (73.7-68.4%), decreasing the stability to certain extent. It is generally accepted that the hydrophobicity is related to the stability of the membrane structure, and thus might attract the interaction between the virus-origin M protein and the host-origin membrane.

3. Substitution leading to the change of hydrophilicity, related to interaction with interior molecules.

The C/T transition in TMII leads to Ala/Val changes. Biochemically, both Ala and Val are both non-polar, non-hydrophilic, and have the same pI (pI 5.6). Therefore, no change is predicted concerning the subregional pI (pI 7.90), hydrophilicity (9.10%), (+) charge (4.50%), (-) charge (0) in TMII by the substitution. The only effect would be the change of hydrophobicity (63.6-68.2%), which would also affect the membrane structure by increasing its stability. The fourth substitution is a T/C transition leading to a Ser/Pro change. Both Ser and Pro are neutral and

have the same pI (pI 5.6), but Ser is absolutely hydrophilic. The only predicted change is a slight alteration in the subregional hydrophilicity (45.4-44.7%), leaving all the other physical and chemical features unaltered. However, hydrophilicity should be important to this interior region if the internal environment is taken into consideration.

Possible origin and evolution of SARS-CoV

The comparative analysis of the genome sequences of 361 ssRNA viruses demonstrates that the M protein is uniquely present in the seventeen currently available coronavirus genomes. The analysis on the published genome sequences of ten members in the family *Coronaviridae* has placed the SARS virus outside the three known groups. According to the unrooted phylogenetic tree we have proposed based on the M protein, the BCoV in Group 3 would be relatively closer to SARS-CoV. This finding for the M protein differs from the phylogenetic analysis for other structural proteins of the SARS-CoV (*6, 14, 15*). However, our data based on the M protein, together with other ORFs and the entire genome, lend strong support to the conclusion that SARS-CoV is a variant in *Coronaviridae* that has been existing in a reservoir somewhere in nature, perhaps in non-human animal(s), for a long time and has only recently found a new pathway to humans. If SARS-CoV has been latent in its original host, it might have recently gained an extraordinarily strong selective advantage to expand its host range, establishing a high virulence to humans, through or after passing into humans. Alternatively, the virus might have passed back and forth from wild animals to humans for some time, lately changing from a wild form to a potent one.

Materials and Methods

Sample processing and sequencing

Samples from the SARS patients, their processing and preparation, as well as sequencing have been described previously (*6*). Briefly, 6 isolates of coronavirus-like virus were isolated from Vero-E6 cell cultures infected with tissues from SARS patients. Virus preparation, RT-PCR and cloning were carried out according to standard protocols at BGI (Beijing Genomics Institute, http://www.genomics.org.cn). Sequencing was performed with MegaBACE 1000 (Amersham, New Jersey, USA).

Sequences and computing analysis

The complete (BJ01 and BJ04) and draft (BJ02, BJ03 and ZJ01) sequences of five isolates generated by BGI and WIGS (James D. Watson Institute of Genome Sciences) and deposited in GenBank (accession numbers: AY278488, AY279354, AY278490, AY278489, AY297028.1) were used for the analysis. Complete sequences of other 12 isolates of SARS-CoV recently published by other institutions, and all other coronaviruses and ssRNA viruses used for this study were downloaded from GenBank (http://www.ncbi.nlm.nih.gov).

The sequences were assembled by using Phrap (http://www.phrap.org). ORFs were identified by ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Composition of the nucleotide sequences was analyzed by DNA_GC_Calculator (http://www.genome.iastate.edu/ftp/share/DNAgcCal/) and hypothesized physical and chemical features of the protein and glycosylation sites were examined by using Compute PI/MW (http://us.expasy.org/tools/pi_tool.html) and NetNGlyc1.0 (http://www.cbs.dtu.dk/services/NetNGlyc) respectively. Comparative analysis was performed using Blast against the non-redundant protein and non-redundant nucleotide sequences databases (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). Multiple sequence alignment was performed with ClustalW1.8 (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalW). TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/), TMMTOP (http://www.enzim.hu/hmmtop/index.html) and Antheprot 5.0 (http://antheprot-pbil.ibcp.fr) were used for transmembrane and secondary structure prediction. Protease cleavage sites were analysed by PeptideCutter (http://ca.expasy.org /tools/peptidecutter/).

All the bioinformatics analyses were performed on supercomputers DOWNING 2000, DOWNING 3000 (DOWNING Computers Inc., Beijing, China), SUN E10K (SUN Microsystems Inc., California, USA), SGI3800 (Silicon Graphics, Inc., California, USA) and IBM P690 (IBM Corp., New York, USA) at BGI.

Synthetic oligopeptides and ELISA

Based on hydropathicity and antigenicity analyses of the M protein, we designed 8 oligopeptides with 16-25 amino acid residues (Table 4). These oligopeptides cover almost all unique regions (69.2%) of SARS-CoV's M protein. Two Lys residues were added in front of the oligopeptides for linkage to a conjugate. Oligopeptides were synthesized by the Hangzhou Zhongtai Company and then characterized by HPLC and mass spectrometry.

Table 4 Synthetic Oligopeptides of the M Protein in SARS-CoV

Name	Amino acid residua	Position (a. a.)
M001	KKMADNGTITVEELKQLLEQWN	1-20
M017	KKEQWNLVIGFLFLAWIMLLQFAYSNR	17-41
M103	KKARTRSMWSFNPETNILLNVPLR	103-124
M119	KKLNVPLRGTIVTRPLMESELVIG	119-140
M137	KKLVIGAVIIRGHLRM AGHSLGR	137-157
M162	KKKDLPKEITVATSRTLSYYKLGASQR	162-186
M189	KKDSGFAAYNRYRIGNYKLNTDHAG	189-211
M206	KKNTDHAGSNDNIALLVQ	206-221

Synthetic oligopeptides were fixed on 96-well plates at 1 μg/mL in a volume of 100 μL for each well. 100 μL Dilution Solution BSA (3%) was added to each well and then mixed with 10 μL serum. After being incubated for 60 min at 37 $^\circ$C and washed three times (5 min each) with PBST (100 mM Tris, 120 mM NaCl, 0.1% Tween-20, pH 7.9), 100 μL enzymes-linked solutions for Goat anti-human IgG (Beijing Zhongshan Company) were added, incubated again for 20 min at 37 $^\circ$C, and then washed. The peroxidase reaction was visualized using the o-pheylenediamine solution as substrate. After 10 min incubation at 37 $^\circ$C, the reaction was stopped by adding 50 μL of 4 M sulphuric acid to each well. Optical density (O.D.) at 450 nm was measured with an automatic ELISA reader.

Acknowledgements

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 indebted to collaborators and clinicians from Peking Union Medical College Hospital, National Center of Disease Control of China, and the Municipal Governments of Beijing and Hangzhou. We thank Dr. Gwendolyn Zahner, Visiting Professor at BGI, for her comments on drafts of this manuscript.

References

1. Michael, M., *et al.* 1986. *Coronaviruses.* pp 108-135. Plenum Press, New York, USA.
2. Lapps, W., *et al.* 1987. Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology.* 157: 47-57.
3. Fleming, J.O., *et al.* 1989. Monoclonal antibodies to the matrix (E1) glycoprotein of mouse hepatitis virus protect mice from encephalitis. *Virology* 168: 162-167.
4. Armstrong, J., *et al.* 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. *Nature* 308: 751-752.
5. Narayanan, K., *et al.* 2000. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J. Virol.* 74: 8127-8134.
6. Qin, E. D., *et al.* 2003. A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01). *Chin. Sci. Bull.* 48: 10941-10948.
7. Melen, K., *et al.* 2003. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* 327: 735-744.
8. Narayanan, K. and Makino, S. 2001. Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J. Virol.* 75: 9059-9067.
9. Kuo, L. and Masters, P.S. 2002. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J. Virol.* 76: 4987-4999.
10. Rodriguez I, *et al.* 2002. Mombaerts P. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat. Neurosci.* 5: 134-140.
11. Lecomte, J., *et al.* 1987. Protection from mouse hepatitis virus type 3-induced acute disease by an anti-nucleoprotein monoclonal antibody. Brief report.

Arch. Virol. 97: 123-130.

12. de Haan, C.A., *et al.* 2002. O-glycosylation of the mouse hepatitis coronavirus membrane protein. *Virus Res.* 82: 77-79.
13. An, S.W., *et al.* 1999. Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer. *J. Viol.* 73: 7853-7859.
14. Marra, M.A., *et al.* 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300: 1399-1404.
15. Rota, P.A., *et al.* 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300: 1394-1399.