research articles Journal of proteome research

Proteomic Analysis of SARS Associated Coronavirus Using Two-Dimensional Liquid Chromatography Mass Spectrometry and One-Dimensional Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis Followed by Mass Spectroemtric Analysis

Rong Zeng,*,† Hong-Qiang Ruan,† Xiao-Sheng Jiang,† Hu Zhou,† Lv Shi,† Lei Zhang,† Quan-Hu Sheng,† Qiang Tu,† Qi-Chang Xia,† and Jia-Rui Wu*,‡

Research Center for Proteome Analysis, Key Lab of Proteomics, Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 YueYang Road, Shanghai, 200031, China

Received November 25, 2003

The proteomes of the severe acute respiratory syndrome-associated coronavirus (SARS-CoV) and its infected Vero E6 cells were detected in the present study. The cytosol and nucleus fractions of virus-infected cells as well as the crude virions were analyzed either by one-dimensional electrophoresis followed by ESI−MS/MS identification or by shotgun strategy with two-dimensional liquid chromatography-ESI−MS/MS. For the first time, all of the four predicted structural proteins of SARS-CoV were identified, including S (Spike), M (Membrane), N (Nucleocapsid), and E (Envolope) proteins. In addition, a novel phosphorylated site of M protein was observed. The combination of these gel-base and non-gel methods provides fast and complimentary approaches to SARS-CoV proteome and can be widely used in the analysis of other viruses.

Keywords: coronavirus • proteome • mass spectrometry • glycosylation • phosphorylation

Introduction

Recently, a novel coronavirus has been identified, which caused the outbreak of severe acute respiratory syndrome (SARS) worldwide.[1,2] The analysis of the complete nucleotide sequences of SARS-associated coronavirus (SARS-CoV) showed that its genome organization was similar to that of other known Coronaviruses.[3,4] The genome of SARS-CoV is approximately 30 kb in size and has 14 predicted open reading frames.

The information of the SARS-CoV genome sequence provides clues for identification of the viral proteins. It looks easy to analyze the entire genome of coronaviruses, but the identification of protein components of coronaviruses has proven to be a difficult task. According to the annotation of its genome and the knowledge about other known coronaviruses, four types of structural proteins of SARS-CoV have been predicted.[5] The spike (S) glycoprotein, together with small envelope (E) protein and matrix (M) glycoprotein, consists of the viral envelope, whereas the nucleocapsid (N) protein interacts with genomic RNA of the virus to form the viral nucleocapsid.[5−6]

Very soon after the SARS-CoV genome sequencing, Krokhin and his colleagues in Canada reported the identification of two major structural proteins, spike glycoprotein and nucleocapsid protein, with mass spectrometry.[7] However, M and E proteins of SARS-CoV have not been reported so far.

In the present study, Vero E6 cells, which are widely used as a cell model for analysis of coronaviruses, were infected with SARS-CoV solution and analyzed with proteomic approaches. By using 2D-LC−MS/MS and 1D-PAGE followed by ESI−MS/MS, we identified all of the four predicted structural proteins from the virus-infected cells. Furthermore, we also identified these four structural proteins from the crude SARS-CoV fraction with the same approaches. In addition, a novel phosphorylated site of M protein was identified.

Materials and Methods

Materials. Chemicals used for gel electrophoresis were from Bio-Rad (Hercules). Formic acid (FA), guanidine hydrochloride were obtained from Sigma (St. Louis,). Acetonitrile (ACN) HPLC grade was from Fisher (Fair Lawn). Trypsin (sequencing grade) and *N*-glycosidase F were obtained from Roche (Mannheim).

Cell Culture and Virus Infection. African green monkey kidney cells (Vero E6, ATCC) were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Gibco-BRL) supplemented with 10% fetal bovine serum (FBS, Gibco-BRL) at 37 °C in a 5% CO_2.

For virus infection, Vero E6 cells were treated with the DMEM medium (2% FBS) containing SARS-CoV virions (BJ-01

* To whom correspondence should be addressed. Dr. Jia-rui Wu, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, CAS, 320 Yue-Yang Road, Shanghai 200031, China. Tel: 86-21-54921128. Fax: 86-21-54921011. E-mail: wujr@sibs.ac.cn. Dr. Rong Zeng, Research Center for Proteome Analysis, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China. Tel: 86-21-54920170. Fax: 86-21-54920171. E-mail: zr@sibs.ac.cn.

† Research Center for Proteome Analysis, Key Lab of Proteomics.

‡ Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology.

10.1021/pr034111j CCC: $27.50

Published on Web 04/08/2004

Table 1. Identified Peptides of Nucleocapsid Protein with ESI−MS/MS

identifed method	peptide sequence[a]	residue position	calculated MH^+
2D-LC−MS/MS	S*DNGPQSNQRSAPR	1−14	1145.12
2D-LC−MS/MS	S*DNGPQSNQRSAPRITFGGPTDST DNNQNGGR	1−32	3389.42
2D-LC−MS/MS	SAPRITFGGPTDSTDNNQNGGR	11−32	2263.33
2D-LC−MS/MS, 1D-PAGE	ITFGGPTDSTDNNQNGGR	15−32	1851.87
2D-LC−MS/MS, 1D-PAGE	ITFGGPTDSTDNNQNGGRNGARPK	15−38	2475.58
2D-LC−MS/MS, 1D-PAGE	RPQGLPNNTASWFTALTQHGK	41−61	2325.57
1D-PAGE	EELRFPR	62−68	947.07
2D-LC−MS/MS, 1D-PAGE	GQGVPINTNSGPDDQIGYYR	69−88	2152.27
1D-PAGE	GQGVPINTNSGPDDQIGYYRR	69−89	2308.45
1D-PAGE	MKELSPR	101−107	861.05
2D-LC−MS/MS, 1D-PAGE	WYFYYLGTGPEASLPYGANK	108−127	2298.54
1D-PAGE	WYFYYLGTGPEASLPYGANKEG IVWVATEGALNTPK	108−143	3965.42
2D-LC−MS/MS, 1D-PAGE	EGIVWVATEGALNTPK	128−143	1685.90
2D-LC−MS/MS, 1D-PAGE	EGIVWVATEGALNTPKDHIGTR	128−149	2365.63
2D-LC−MS/MS, 1D-PAGE	DHIGTRNPNNNAATVLQLPQGTTLPK	144−169	2772.07
2D-LC−MS/MS, 1D-PAGE	NPNNNAATVLQLPQGTTLPK	150−169	2092.34
2D-LC−MS/MS, 1D-PAGE	GFYAEGSR	170−177	886.93
2D-LC−MS/MS	GFYAEGSRGGSQASSRSSSRSR	170−191	2278.35
2D-LC−MS/MS	GFYAEGSRGGSQASSR	170−185	1617.66
2D-LC−MS/MS	GFYAEGSRGGSQASSRSSSR	170−189	2035.08
2D-LC−MS/MS	GNSRNSTPGSSRGNSPAR	192−209	1802.85
2D-LC−MS/MS, 1D-PAGE	MASGGGETALALLLLDR	220−236	1688.97
2D-LC−MS/MS, 1D-PAGE	MASGGGETALALLLLDRLNQLESK	220−243	2501.89
1D-PAGE	LNQLESK	227−233	831.94
1D-PAGE	VSGKGQQQQGQTVTK	235−249	1574.72
1D-PAGE	VSGKGQQQQGQTVTKK	234−249	1702.89
2D-LC−MS/MS	TATKQYNVTQAFGR	263−276	1585.75
2D-LC−MS/MS, 1D-PAGE	K.QYNVTQAFGR.R	267−276	1184.29
2D-LC−MS/MS, 1D-PAGE	RGPEQTQGNFGDQDLIR	277−293	1932.04
2D-LC−MS/MS	RGPEQTQGNFGDQDLIRQGTDYK	277−299	2624.77
2D-LC−MS/MS, 1D-PAGE	GPEQTQGNFGDQDLIR	278−293	1775.86
2D-LC−MS/MS	HWPQIAQFAPSASAFFGMSR	300−319	2237.53
2D-LC−MS/MS, 1D-PAGE	IGMEVTPSGTWLTYHGAIK	320−338	2062.38
2D-LC−MS/MS	IGMEVTPSGTWLTYHGAIKLDDK	320−342	2533.89
2D-LC−MS/MS, 1D-PAGE	LDDKDPQFK	339−347	1106.21
2D-LC−MS/MS, 1D-PAGE	LDDKDPQFKDNVILLNK	339−355	2016.28
2D-LC−MS/MS	DPQFKDNVILLNK	343−361	1544.78
2D-LC−MS/MS, 1D-PAGE	DNVILLNK	348−355	929.10
2D-LC−MS/MS	DNVILLNKHIDAYKTFPPTEPK	348−369	2554.93
2D-LC−MS/MS, 1D-PAGE	KKTDEAQPLPQR	374−385	1411.59
2D-LC−MS/MS, 1D-PAGE	KTDEAQPLPQR	375−385	1283.42
2D-LC−MS/MS, 1D-PAGE	TDEAQPLPQR	376−385	1155.24
1D-PAGE	QKKQPTVTLLPAADMDDFSR	386-405	2262.57
2D-LC−MS/MS	KQPTVTLLPAADMDDFSR	388−405	2006.27
2D-LC−MS/MS	KQPTVTLLPAADMDDFSRQ LQNSMSGASADSTQA	388−421	3583.91
2D-LC−MS/MS, 1D-PAGE	QPTVTLLPAADMDDFSR	389−405	1878.10
2D-LC−MS/MS, 1D-PAGE	QLQNSMSGASADSTQA	406−421	1596.66

[a] Asterisk indicates acetylation.

isolate, provided by Academy of Military Medical Sciences) for 1 h, of which $TCID_{50}$ (tissue culture infectious dose) was identified as 10^6 dilution. The virus-medium was removed after the infection, and the infected cells were cultured in the DMEM medium with 2% FBS at 37 °C in a 5% CO_2. All of the experiments using the virus were carried on in Bio-safety Level 3 laboratory.

Collection of Cytosol and Nuclear Fractions of Infected Cells. According to Hasbold et al. with minor modifications,[8] Vero E6 cells were infected with SARS-CoV virions for 24 h, of which no cell-lyses was observed by microscopy. The infected cells then were washed with cold phosphate-buffer two times and incubated with a solution containing 40 mM Tris (pH 8.3) and 0.5% Nonident P-40 at room temperature for 5 min. The cell lysate was collected and centrifuged at 8000 rpm for 5 min. After the centrifugation, the supernatant was collected and heated at 100 °C for 5 min as cytosol fractions, while the pellet was resuspended with reducing loading buffer (50 mM Tris, pH 6.8, 2% SDS, 10% glycerol, 100 mM DTT, 0.1% bromophenol blue) and heated at 100 °C for 5 min as nuclear fractions.

Collection of crude SARS-CoV virions in medium. After 48 h post-infection, more than 80% of infected Vero E6 cells were lysed by the virus. The medium containing virus particles was collected and centrifuged at 12 000 rpm for 30 min to remove the cell debris. Then the supernatant was centrifuged with microcon tubes (Millipore, YM-100) and the up-solution in the microcon tube was collected as crude SARS-CoV virions.

One-Dimensional SDS Electrophoresis (1D-SDS-PAGE). Either the cytosol and nucleus fractions of infected Vero E6 cells, or the crude virus in medium were mixed with the equal volume of denaturing buffer (100mM Tris, 1% SDS) and boiled for 10 min. The mixtures were subjected to SDS-PAGE with 7.5−17% gradient gel.

Tryptic Digestion of In-Gel Proteins. The interested gel pieces were cut from the gels and destaied twice with 100 mM NH_4HCO_3 and 30% acetonitrile, and washed with water. These

Table 2. Identified Peptides of Spike Protein with ESI−MS/MS. (A) Indicates the Identified Peptides before De-glycosylation and (B) Presents Additional Peptides Identified after De-glycosylation[a]

identified method	peptide sequence	residue position	calculated MH^+
	A		
2D-LC−MS/MS, 1D-PAGE	DLPSGFNTLKPIFK	208−221	1577.85
2D-LC−MS/MS	SFEIDKGIYQTSNFR	292−306	1805.97
2D-LC−MS/MS	LNDLCFSNVYADSFVVK	374−390	1992.21
2D-LC−MS/MS	LNDLCFSNVYADSFVVKGDDVR	374−395	2534.76
2D-LC−MS/MS, 1D-PAGE	QIAPGQTGVIADYNYK.L	396−411	1738.92
2D-LC−MS/MS, 1D-PAGE	NIDATSTGNYNYK	427−439	1461.52
2D-LC−MS/MS	DISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYR	454−495	4845.28
2D-LC−MS/MS	VVVLSFELLNAPATVCGPK	496−514	2015.38
2D-LC−MS/MS	NQCVNFNFNGLTGTGVLTPSSK	522−544	2356.57
2D-LC−MS/MS	FQPFQQFGR	545−553	1155.29
2D-LC−MS/MS, 1D-PAGE	DVSDFTDSVRDPK	554−566	1481.55
1D-PAGE	ALSGIAAEQDR	748−758	1131.22
2D-LC−MS/MS, 1D-PAGE	ALSGIAAEQDRNTR	748−761	1502.62
2D-LC−MS/MS	ALSGIAAEQDRNTREVFAQVK	748−768	2304.55
2D-LC−MS/MS	EVFAQVK	762−768	820.96
2D-LC−MS/MS	RSFIEDLLFNK	797−807	1382.59
2D-LC−MS/MS, 1D-PAGE	QYGECLGDINAR	818−836	1396.48
2D-LC−MS/MS, 1D-PAGE	FNGIGVTQNVLYENQK	888−903	1825.02
2D-LC−MS/MS, 1D-PAGE	AISQIQESLTTTSTALGK	912−929	1850.06
2D-LC−MS/MS	AISQIQESLTTTSTALGKLQDVVNQNAQALNTLVK	912−946	3700.15
2D-LC−MS/MS, 1D-PAGE	LQDVVNQNAQALNTLVK	930−946	1869.11
2D-LC−MS/MS	QLSSNFGAISSVLNDILSR	947−965	2022.25
2D-LC−MS/MS, 1D-PAGE	LQSLQTYVTQQLIR	978−996	1691.95
2D-LC−MS/MS	MSECVLGQSK	1011−1020	1139.3
1D-PAGE	EELDKYFKNHTSPDVDLGDISGINASVVNIQK	1132−1163	3547.87
1D-PAGE	EIDRLNEVAK	1164−1173	1187.33
2D-LC−MS/MS	FDEDDSEPVLK	1238−1248	1294.35
	B deglycopeptides found after PNGase F treatment		
1D-PAGE, deglycosylation	LPLG*I*N*I*TNFR	222−232	1258.50
1D-PAGE, deglycosylation	YDE*N*GTITDAVDCSQNPLAELK	266−287	2454.58
1D-PAGE, deglycosylation	FP*N*ITNLCPFGEVF*N*ATK	316−333	2070.33
1D-PAGE, deglycosylation	EGVFVF*N*GTSWFITQR	1074−1089	1889.10
1D-PAGE, deglycosylation	NL*N*ESLIDLQELGK	1174−1187	1586.77

[a] Italic N indicates potential *N*-glycosylation site.

Table 3. Identified Proteins of Membrane Protein with ESI−MS/MS

identified method	peptide sequence	residue position	calculated MH^+
2D-LC−MS/MS	QLLEQWNLVIGFLFLAWIML LQFAYSNR	14−41	3429.08
2D-LC−MS/MS	SMWSFNPETNILLNVPLR	107−124	2132.47
2D-LC−MS/MS	CDIKDLPK	158−165	989.14
2D-LC−MS/MS, 1D-PAGE	VGTDSGFAAYNR	186−197	1258.32
2D-LC−MS/MS	VGTDSGFAAYNRYRIGNYK	186−204	2153.34
2D-LC−MS/MS, 1D-PAGE	LNTDHAGSNDNIALLVQ	205−221	1795.93

gel pieces were incubated with 100 mM NH_4HCO_3 containing 10mM DTT at 56 °C for 30 min, and then incubated with 60 mM iodoacetamide at room temperature for 20 min. Gel pieces were then dehydrated in 100 μL of 100% acetonitrile. 12.5 ng/μL trypsin (Sequencing grade, Promega,) was added to cover the gel pieces and incubated at 37 °C overnight The gel pieces were then extracted twice in 100 μL of 60% acetonitrile, 0.1% trifluoroacetic acid (Fluka) with ultrasonication for 10 min. The supernatants were pooled and lyophilized in a SpeedVac for mass spectrometric analysis.

Tryptic Digestion of Protein Mixture. The cell lysate or the crude virus fraction was reduced with 10mM DTT at 37 °C for 4 h, and then alkylated with 60 mM iodoacetamide at room temperature for 30 min. The protein buffer was exchanged to digestion buffer (100 mM ammonium bicarbonate, pH8.5) and incubated with trypsin at 37 °C for 24 h.

***N*-Glycosidase F Deglycosylation of S Protein.** 1 unit of *N*-glycosidase F in 4 μL H_2O was added to the peptide digests of in-gel S-protein dissolved in 100 mM NH_4HCO_3 to a concentration of 1 mg/ml, pH8.3). The mixture was incubated at 37 °C for overnight.

1D- and 2D-LC−ESI−MS/MS. For in-gel protein identifications, 1D-LC−ESI−MS/MS (LCQ Deca XP Plus Thermo Finnigan) was used. Peptides were separated by reverse-phase chromatography using a 0.18 mm × 100 mm column (BioBasic-C18, Thermo Hypersil-Keystone) at a flow rate of 2 μL after splitting. Protein digests of whole protein mixture were analyzed with 2D-LC−MS/MS system (ProteomeX, Thermo Finnigan). The first dimensional was strong cation exchange (Biobasic-SCX; 0.32 mm × 100 mm, Thermo Hypersil-Keystone). The elution gradients were 0, 25, 50, 75, 100, 150, 200, 400, and 800 mM ammonium chloride. The second dimension was reversed phase as used in 1D-LC−MS/MS.

The MS spray voltage was maintained at 3.3 KV, and the temperature of ion transfer tube was at 150 °C. The collision energy of MS/MS was 35%. Each scan event was composed of one full scan MS and three MS/MS of the most intensive peaks. Dynamic exclusion was also applied.

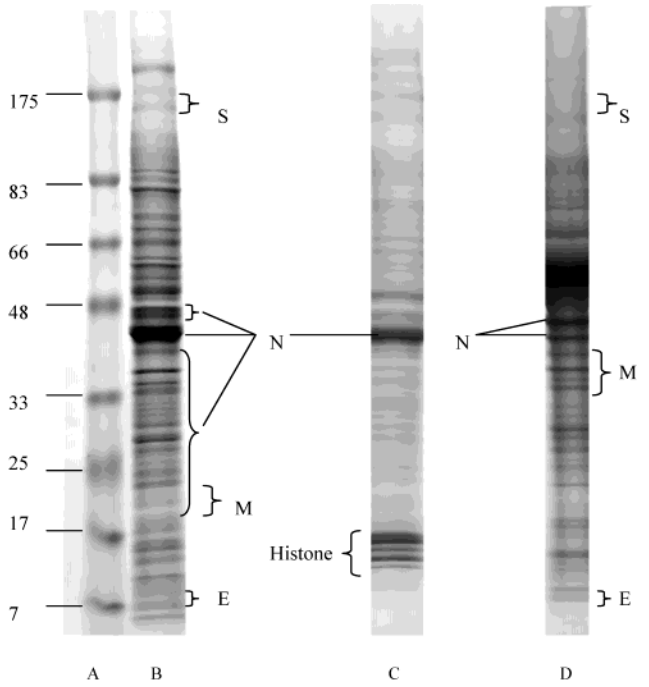

Figure 1. 1D-gel maps of SARS-CoV and infected Vero E6 cells. Lane A is the molecular markers. Lane B is the cytosol fraction of E6 cells infected with SARS-CoV after 24 h. Lane C is the nucleus fraction of E6 cells infected with SARS-CoV after 24 h. Lane D is the crude SARS−CoV virus fraction.

Data Analysis. Protein identification was performed with BioWorks version 3.1(Thermo Finnigan,) and SEQUEST algorithm. Since Vero E6 was derived from monkey, both the human-database and the SARS-database from NCBI were merged. The MS results were analyzed against either the merged database or SARS-database alone. The analyzed data were further filtered with Xcorr (1 + > = 1.8, 2 + > = 2.0, 3+ > = 2.5).

Results

Identification of SARS-CoV Structural Proteins with Two Complimentary Proteomic Approaches. When we obtained the virus-infected cells and crude virions, the first step was to analyze the protein mixture with shotgun strategy using 2D-LC−MS/MS, which is the most faster and straightforward means to detect what kinds of the viral proteins in these mixtures. M, S, and N proteins were identified from the whole lysate of virus-infected cells and crude virion solution (Tables 1−3), while E protein was not identified with 2D-LC−MS/MS.

On the other hand, the traditional way for identification of proteins, one-dimensional electrophoresis followed by ESI−MS/MS, were also applied. In the present study, the cytosol and nuclear fractions as well as crude virions were subjected to 1D-PAGE (Figure 1, Lanes B, C, and D). The interested gel-bands were cut out and then analyzed by 1D-LC−ESI−MS/MS. The results showed the identification of these four predicted structural proteins either from the cytosol of infected cells or from the crude SARS-CoV virions (Figure 1, Lanes B, C, and D, Figure 4; also see Tables 1−3). And interestingly, a novel phosphorylated site of M protein was identified by this method.

Identification of Nucleocapsid (N) Protein. The coronavirus nucleocapsid (N) protein is the most abundant virus-derived protein produced throughout the process of the virus infection. It was easily to identify N protein using either 1D-PAGE followed by ESI−MS/MS or 2D-LC−MS/MS (Table 1) By 2D-LC−MS/MS, the sequence coverage reaches 85.03%, while 1D-PAGE−MS/MS gets 68.41% of the sequence., It was showed that the nonredundant protein coverage reached 89.54% according to the MS/MS of peptides. In addition, N protein displayed multiple bands below the major band of 48 KD (Figure 1, Lane B), suggesting the degradation of N protein. It was observed that N protein presented different major isoforms in the infected cells and the virions 48 KD band was the dominant component in the virions, whereas the band at 46 KD becomed the major one in the cytosol fraction of the infected cells (Figure 1, Lanes B and D). Interestingly, N protein was also found in nucleus fraction of the infected cells, in which 46 KD band was dominant (Figure 1, Lane C).

It was reported that N protein existed as phosphorylated forms in mature viral particles.[9] When N protein entered the host cells during the process of the virus infection, it would be de-phophorylated and this de-phosphorylated form could enter the nucleus to affect the gene transcription of the host cells.[9] Our preliminary work supported the previous report and bioinformatics prediction.

Identification of Spike (S) Protein. S proteins as a major structural protein of SARS-CoV locate on the surface of viral particles. Our present works showed that S proteins were detected in both the infected cells and the crude virus fraction with 1D-PAGE followed with ESI−MS/MS or with 2D-LC−MS/MS. The combination with these two kinds of proteomic approaches provided the coverage of 30.19% amino acids of S protein. In the 1D-PAGE, S protein appears at 175 KD region (Figure 1, Lanes B and D). The in-gel protein digests of S-protein was treated by PNGase to remove the *N*-glycosylation. 5 more peptides with potential *N*-glycosylated sites were identified, contributing additional 6.45% (Table 2) of the identification coverage, and the total coverage reached 36.65%.

Characterization of Membrane (M) Protein. The membrane protein should be also on the surface of virions and coupled with S protein. Table 3 lists the identified 6 peptides of M protein by MS/MS, and the protein coverage was 50.68%. 2D-LC−MS/MS identified all of the 6 peptides while 1D-PAGE−MS/MS only obtained 2 peptides (Table 3). The M protein is composed with 221 amino acids with theoretical molecular weight of 25 KD. M protein is thought as a glycoprotein with higher molecular weight than theoretical value.[4] Indeed, M proteins in the crude virus fraction were observed in the regions of 33−42 KD, while M proteins in the infected cells were detected only in the region of 18−23 KD (Figure 1, compare Lanes B and D), which may indicate the modifications occurring on mature M proteins in the virions. Interestingly, we identified a form of phosphorylated M proteins from the crude virus fraction by the MS/MS (Figure 2), although no M glycoprotein was found in the present study. The site of phosphorylation was located at the C terminus of M protein (Figure 2). The complete and continuous appearance of b ions strongly supported the existence of phosphorylatded peptides and the neutral loss of phosphorylatded peptide in its MS/MS spectrum. The results indicate that M protein may be a phosphoprotein, while the function of phosphorylation of M protein remains to be uncovered.

Analysis of Envolope (E) Protein. From the annotation of genome sequence, it was predicted the SARS-CoV has a small envelope protein on its surface.[3,4] However, the identification

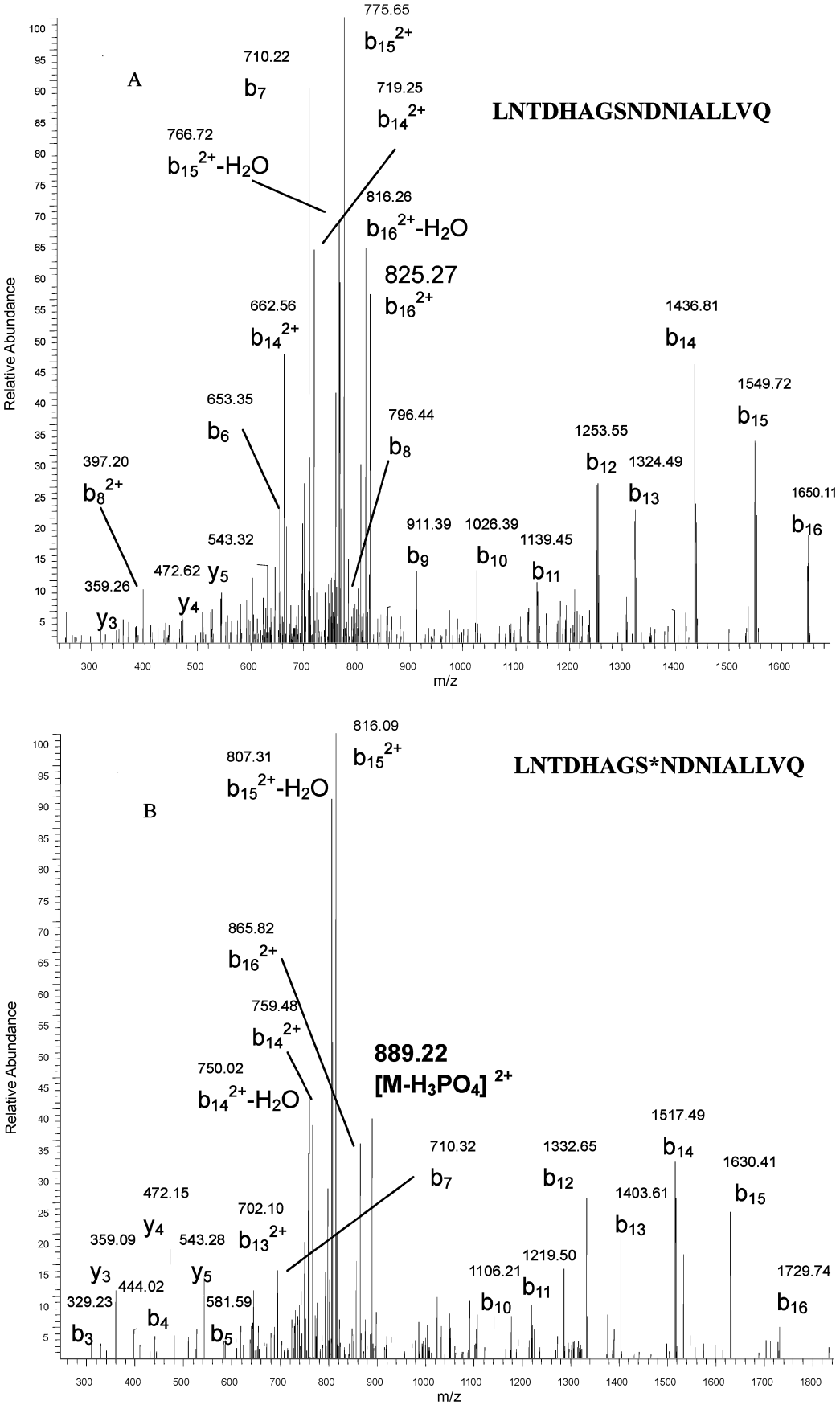

Figure 2. MS/MS spectra of C-terminal peptides (LNTDHAGSNDNIALLVQ) of M protein from the crude virus fraction. A shows the doubly charged unphosphoyylated peptide (*m*/*z* **898.09**). B shows the doubly charged phosphorylated peptide (*m*/*z* **937.96**). S* indicates the phosphorylated Ser, and ion at *m*/*z* 889.2 indicates the ion with neutral loss of H_3PO_4.

MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTAL**R**LCAYCCNIVNVSLV**K**
PTVYVYS**R**V**K**NLNSSEGVPDLLV

Figure 3. Protein sequence of E protein, tryptic cleavage sites are bold and identified peptide is underlined.

of E protein of coronaviruses has been thought a difficult task due to its properties. First, E protein is low-abundant in the family of coronaviruses.[4] Second, the analysis of E-protein sequence showed only four tryptic cleavage sites (R38, K53, R61, and K63, shown in Figure 3). The site of K53 is just before a proline which may prevent the cleavage with trypsin. Third, E protein contains three cysteines, which indicates that E protein may form disulfide bonds within itself or with other proteins, making it difficult to be reduced and digested. In addition, E protein is a very hydrophobic protein because the amino acids 17−34 are predicted to be embedded in the viral membrane.

In the present work, we failed to obtain E protein with 2D-LC−MS/MS. However, one peptide of C-terminal of E protein was eventually identified either in the cytosol of the infected cells and the crude virus fraction separated by 1D-PAGE (Figure 1, Lanes B and D). Figure 4 showed the MS/MS spectrum of C-terminal peptide, VKNLNSSEGVPDLLV, with good quality of the ion signals and intensive peaks of y_5 and b_{10}, which was from the easy fragmentation at N-terminus of proline residue. The results indicate that E protein is expressed in SARS-CoV virus but with very low abundance.

Discussion

To our knowledge, the present work first time showed identification of all of the four structural proteins of SARS-CoV, spike, membrane, nucleocapsid, and envelope proteins, on the protein level. Moreover, the combination of 1D-SDS-PAGE followed by ESI−MS/MS and 2D-LC−MS/MS proved to be an efficient and complimentary way to identify viral proteins. On one hand, the shotgun method seems to get more identification coverage than 1D-PAGE followed by MS/MS, and it is also much faster than gel-based assay. On the other hand, comparing to the using mild denaturing condition to maitain the trypsin activity in the shotgun approach, the proteins could be strongly denatured during the step of 1D-PAGE prior to in-gel tryptic digestion, which is very helpful for tryptic digestion of the very hydrophobic proteins such as E protein. The gel-base method is advantageous in resolving different components of proteins and acquiring detailed information of viral proteins such as locations and modifications. We observed the different compositions of nucleocapsid protein in the cytosol and nucleus fractions of the virus-infected cells and the virions, indicating that nucleocapsid protein has multiple isoforms during the process of infecting the host cells. The N protein was observed in the infected nucleus fraction, consistent with the previous report, which indicates the N protein can enter in the nucleus and affect the gene regulation and cell cycle of the host cells.[10−11]

Membrane protein are composed of 221 amino acids with three transmembrane regions across the viral membrane.[4] The N-terminus of M protein is predicted to be exposed on the surface of virus and the C-terminal region is located inside the virus. The M proteins were found with *O*-glycosylation or *N*-glycosylation in coronavirus family.[12,13] In this study, we first time report the phosphorylated Ser212 at the C-terminus of membrane protein. Using the software NetPho for the predic-

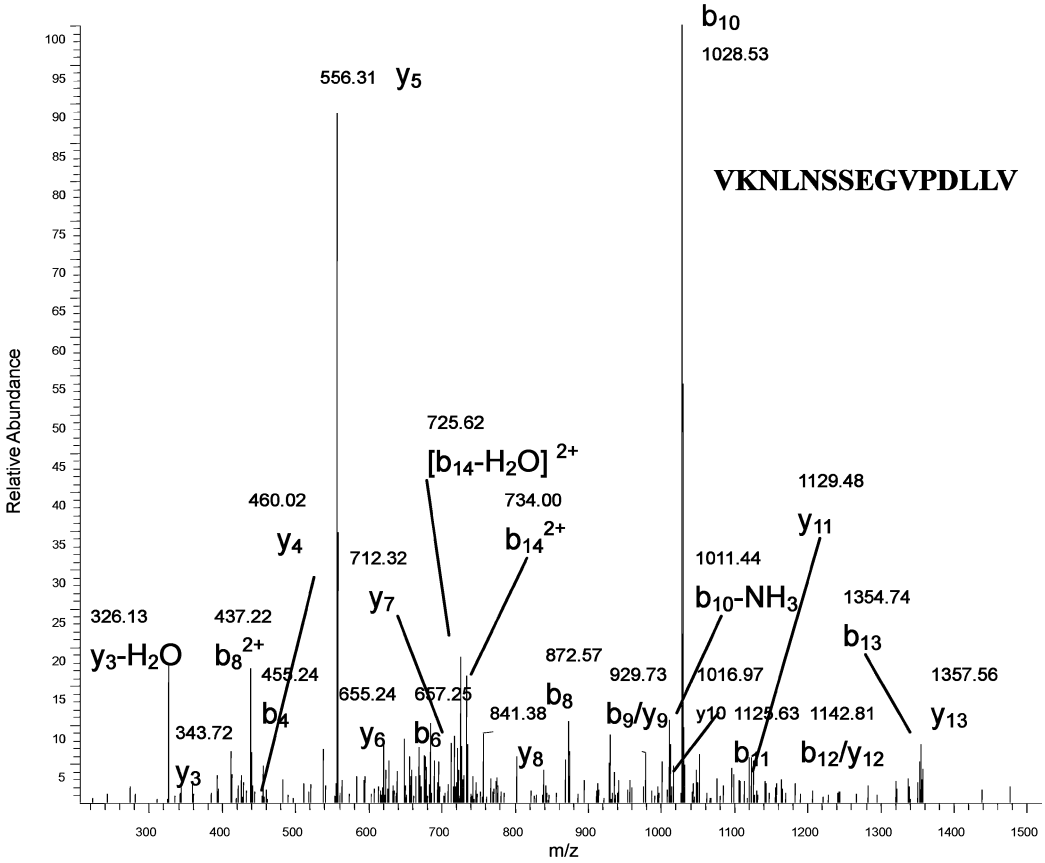

Figure 4. The MS/MS spectra of doubly charged peptide VKNLNSSEGVPDLLV (*m*/*z* **792.80**) from small envelope protein (E protein).

tion of phosporylation-sites (www.cbs.dtu.dk/services), Ser212 of C-terminuse of M protein was predicted as a potential site of phosphorylation, whereas there is no obvious *O*-glycosylation site and only one *N*-glycosylation site at the N-terminus by the prediction(www.cbs.dtu.dk/services). The C-terminus of M protein was reported to be crucial to the assembly of viral envelope and the deletion of single amino acid in this region would be fatal in mouse hepatitis virus.[14,15] It should be interesting in analyzing the biological function of the phosphorylated C-terminus of M protein in SARS-CoV.

The small envelope protein is a transmembrane protein across the membrane.[16,17] According to the genome annotation of SARS-CoV, its E protein contains 76 amino acid, in which 17−34 is transmembrane region and the C-terminus is on the surface of the virus.[3] The E protein in coronavirus was reported to be involved in virus assembly but may have different roles in virus replication in different viruses.[18,19] However, the E proteins of coronavirus are very low abundance compared to N, S, and M proteins,[4,14,16−17] as well as highly hydrophobic. The grand average hydrophobicity (GRAVY) values scores provide an image of the hydrophobicity of the whole protein, usually varying in the range of ±2. Positive score indicates hydrophobic and negative score indicates hydrophilic.[20] The GRAVY of the E protein is 1.141 according to ProtParam (www.expasy.ch), which indicates E protein is very hydrophobic thus difficult to be soluble in lysis buffer. In this study, we first detected the C-terminal tryptic peptide of E protein with MS/MS, confirming the existing of E protein in SARS-CoV.

Conclusions

In summary, we used two complimentary methods, 2D-LC−MS/MS and 1D-PAGE followed by ESI−MS/MS, to analyze the proteins of SARS-CoV. For the first time, we identified all of the four structural proteins, especially the very low-abundant E protein. In addition, different isoforms of N protein and phosphorylated M protein were further identified. The 1D-PAGE gel-based assay can give more information on protein isoforms caused by modification or degradation, while it is time-consuming. 2D-LC−MS/MS makes contribution to rapidly and accurately characterize whether the cells contain virus and obtain most of the identification coverage, which may be used for rapid screening the virus, virus-infected cells or even body fluids containing viruses as a potential diagnostic tool.

Abbreviations: SARS-CoV, severe acute respiratory syndrome associated coronavirus; 1D-PAGE, one-dimensional polyacrylamide gel electrophoresis; 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; 2D-LC, two-dimensional liquid chromatography; MS, mass spectrometry; MS/MS, tandem mass spectrometry; ESI, electrospray; PNGase, F–*N*-Glycosidase F.

Acknowledgment. This work was supported by grants from National Natural Science Foundation #30340031, Chinese Academy of Sciences #KSCX2-SW-225, Science and Technology Commission of Shanghai Municipality #03DZ19119 to J.R.Wu; grants of National High Technology #2001AA233031 and #2002BA711A11to R. Zeng.

References

(1) Peiris, J. S.; Lai, S, T.; Poon, L. L.; et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet.* **2003**, 361: 1319−1325.
(2) Kuiken T.; Fouchier R. A.; Schutten M.; et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet.* **2003**, *362*, 263−270.
(3) Marra, M. A.; Jones, S. J.; Astell, C. R. The Genome sequence of the SARS-associated coronavirus. *Science* **2003**, *300*, 1399−1404.
(4) Rota, P. A.; Oberste, M. S.; Monroe, S. S.; et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *300*, 1394−1399.
(5) Lai, M.; Holmes, K. Coronaviridae: The Viruses and Their Replication. In *Fields Virology*; Knipe, D., et al., Eds.; Lippincott-Willams and Wikins: Philadelphia, 2001, pp 1163−1185.
(6) Holmes, K. V.; Enjuanes, L. The SARS coronavirus: a postgenomic era. *Science* **2003**, *300*, 1377−1378.
(7) Krokhin, O.; Li, Y.; Andonov, A.; Feldmann, H.; et al. Mass Spectrometric Characterization of Proteins from the SARS Virus: A Preliminary Report. Mol Cell Proteomics. **2003**, *2*, 346−356.
(8) Hasbold, J.; Hodgkin, P. D. Flow cytometric cell division tracking using nuclei. *Cytometry* **2000**, **40**, 230−237.
(9) Kalicharran, K.; Dales, S. Dephosphorylation of the nucleocapsid protein of inoculum JHMV may be essential for initiating replication. *Adv. Exp. Med. Biol.* **1995**, *380*, 485−489.
(10) Wurm, T.; Chen, H.; Hodgson, T.; Britton, P.; Brooks, G.; Hiscox, J. A. Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. *J. Virol.* **2001**, *75*, 9345−9356.
(11) Chen, H.; Wurm, T.; Britton, P.; Brooks, G.; Hiscox, J. A. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *Virol.* **2002**, **76**, 5233−5250.
(12) Yamada, Y. K.; Yabe, M.; Ohtsuki, T.; Taguchi, F. Unique N-linked glycosylation of murine coronavirus MHV-2 membrane protein at the conserved O-linked glycosylation site. *Virus Res.* **2000**, *66*, 149−154.
(13) Wit, M.; de Vries, A. A.; Nilsson, T.; Vennema, H.; Rottier, P. J. Structural requirements for *O*-glycosylation of the mouse hepatitis virus membrane protein. *J. Biol. Chem*. **1998**, *273*, 29905−29914.
(14) de Haan, C. A.; Kuo, L.; Masters, P. S.; Vennema, H.; Rottier, P. J. Coronavirus particle assembly: primary structure requirements of the membrane protein. *J. Virol.* **1998**, *72*, 6838−6850.
(15) Kuo, L.; Masters, P. S. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J. Virol.* **2002**, *76*, 4987−4999.
(16) Yu, X.; Bi, W.; Weiss, S. R.; Leibowitz, J. L. Mouse hepatitis virus gene 5b protein is a new virion envelope protein. *Virology* **1994**, *202*, 1018−1023.
(17) Fischer, F.; Stegen, C. F.; Masters, P. S.; Samsonoff, W. A. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J. Virol.* **1998**, *72*, 7885−7894.
(18) Kuo, L.; Masters, P. S. The small envelope protein E is not essential for murine coronavirus replication. *J. Virol.* **2003**, *77*, 4597−4608.
(19) Ortego, J.; Escors, D.; Laude, H.; Enjuanes, L. Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. *J. Virol.* **2002**, *76*, 11 518−11 529.
(20) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105−132.

PR034111J