Antigenicity Analysis of Different Regions of the Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein

ZELIANG CHEN, DECUI PEI, LINGXIAO JIANG, YAJUN SONG, JIN WANG, HONGXIA WANG,
DONGSHENG ZHOU, JUNHUI ZHAI, ZONGMIN DU, BEI LI, MAOFENG QIU, YANPING HAN,
ZHAOBIAO GUO, and RUIFU YANG*

Background: **The widespread threat of severe acute respiratory syndrome (SARS) to human health has made urgent the development of fast and accurate analytical methods for its early diagnosis and a safe and efficient antiviral vaccine for preventive use. For this purpose, we investigated the antigenicity of different regions of the SARS coronavirus (SARS-CoV) nucleocapsid (N) protein.**

Methods: **The cDNA for full-length N protein and its various regions from the SARS-CoV was cloned and expressed in *Escherichia coli*. After purification, all of the protein fragments were printed on glass slides to fabricate a protein microarray and then probed with the sera from SARS patients to determine the reactivity of these protein fragments.**

Results: **The full-length protein and two other fragments reacted with all 52 sera tested. Four important regions with possible epitopes were identified and named as EP1 (amino acids 51–71), EP2 (134–208), EP3 (249–273), and EP4 (349–422), respectively. EP2 and EP4 possessed linear epitopes, whereas EP1 and EP2 were able to form conformational epitopes that could react with most (>80%) of the tested sera. EP3 and EP4 also formed conformational epitopes, and antibodies against these epitopes existed in all 52 of the sera tested.**

Conclusion: **The N protein is a highly immunogenic protein of the SARS-CoV. Conformational epitopes are important for this protein, and antigenicity of the COOH terminus is higher than that of the NH$_2$ terminus. The N protein is a potential diagnostic antigen and vaccine candidate for SARS-CoV.**

© 2004 American Association for Clinical Chemistry

The outbreak in Guangdong Province of China of atypical pneumonia, termed "severe acute respiratory syndrome" (SARS)[1] by the WHO, spread rapidly to many countries during 2003. Thousands of people were infected, and hundreds of them died from this disease *(1–5)*. Our institute was one of the groups that pioneered the isolation and identification of a novel coronavirus (CoV) as the causal agent *(6)*.

CoVs are members of an enveloped virus family that replicate in the cytoplasm of their animal host cells. They have a single-stranded plus-sense RNA genome ~30 kb in length with a 5′-cap structure and a 3′-polyadenylated tail *(7, 8)*. The open reading frame at the 5′ end of the viral genome encodes several nonstructural proteins responsible for replicating the viral genome as well as for generating nested transcripts *(9, 10)*. The 3′ end encodes four structural proteins: spike glycoprotein (S protein), membrane protein (M protein), envelope glycoprotein (E protein), and nucleocapsid protein (N protein). The N protein is a major protein, and its primary function is to form ribonucleoprotein complexes during virion assembling; it has also been proposed to be a multifunctional protein with potential roles in replication, transcription, and translation *(8, 9, 11)*. For the SARS-CoV, the N protein is encoded by an open reading frame located at the 3′ end of the viral genome, and its presence in the virion has been confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry *(5, 12)*.

Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing 100071, China.

*Address correspondence to this author at: Institute of Microbiology and Epidemiology, Academy of Military Medical Science, 20, Dongdajie, Fengtai District, Beijing 100071, People's Republic of China. Fax 86-10-83820748; e-mail yangrf@nic.bmi.ac.cn.

[1] Nonstandard abbreviations: SARS, severe acute respiratory syndrome; CoV, coronavirus; PLA, People's Liberation Army; PBS, phosphate-buffered saline; and FI, fluorescence intensity.

SARS is a new disease to humans, and many aspects of this disease are still unknown, posing a considerable threat to public health. Therefore, accurate diagnosis and efficient vaccines are vital for controlling this disease. The N protein is a relatively conservative and highly immunogenic protein and could be used as a diagnostic antigen and vaccine candidate *(13)*. Shi et al. *(14)* used an N-protein-based ELISA to analyze its antibodies in patient sera and found that antibodies appeared 6 days after the onset of illness. However, the locations of the antigenicity determinants were still unknown. Wang et al. *(15)* analyzed the antigenicity of the N protein by use of synthetic peptides. They identified three highly immunogenic epitopes but found it difficult to identify the conformational epitopes by use of synthetic peptides. Protein fragments seem to be a useful tool for the analysis of conformational epitopes. In the present study, the cDNA for the full-length N protein and its various regions was cloned and expressed in *Escherichia coli*, and the proteins were purified and used to fabricate microarrays that were probed with sera from SARS patients. Four regions important for antigenicity were proposed after detailed interpretation of the results.

Materials and Methods

COLLECTION OF SERA FROM SARS PATIENTS

The sera used in this study were collected from the 301 Hospital of the People's Liberation Army (PLA), 302 Hospital of PLA, 309 Hospital of PLA, Chaoyang Hospital, Ditan Hospital, and Youan Hospital during the SARS outbreak in Beijing from April to June 2003. A total of 625 sera were screened by SARS-CoV lysate-based ELISA to determine the IgG titer against the SARS-CoV, and 52 of these sera, with IgG readouts ranging from 0.30 to 1.10, were used for immunoblotting and protein microarray assays. Eight sera from healthy people without antibodies to the SARS-CoV were used as negative controls.

SEQUENCE ANALYSIS AND PRIMER DESIGN

The sequence of the gene encoding N protein from the SARS-CoV (strain BJ01) was retrieved from GenBank (accession no. AY278488) and analyzed for putative epitopes with the DNASTAR software package (DNAS-TAR, Inc.; http://www.dnastar.com/). On the basis of the predicted epitope locations, 33 primers were designed and synthesized to amplify the fragments spanning these epitopes (Table 1). The designed primers were synthesized by SUNBIO Tech Company.

CLONING, PRODUCTION, AND PURIFICATION OF FULL-LENGTH N PROTEIN AND FRAGMENTS

The SARS-CoV (BJ01) was cultured in Vero E6 cells at 37 °C for 48 h; the medium from the infected cells was then collected for RNA extraction by the QIAamp viral RNA Mini Kit (QIAGEN, Inc.). From the RNA template, single-stranded cDNA was synthesized with use of the primer NO-R (Table 1) and the SuperScript cDNA system

Table 1. Oligonucleotides designed and used in this study.

Name[a]	Position,[b] 5′ end	Primer sequence,[c] 5′–3′
NA-F1	52	AGCTGGATCCTTTGGTGGACCCACAG
NA-F2	91	AGCTGGATCCGGAGGACGCAATGG
NA-F3	160	AGCTGGATCCTTCACAGCTCTCAC
NA-F4	214	AGCTGGATCCGGCGTTCCAATC
NA-F5	403	AGCTGGATCCGCAACTGAGGGAGC
NA-F6	538	AGCTGGATCCGGCAGTCAAGCCTC
NA-F7	667	AGCTGGATCCTTGCTGCTAGACAG
NA-F8	745	AGCTGGATCCAAGAAATCTGCTGC
NA-F9	817	AGCTGGATCCCAAGCATTTGGGAG
NA-F10	874	AGCTGGATCCCTAATCAGACAAGG
NA-F11	1006	AGCTGGATCCGGAGCCATTAAATTG
NA-F12	1058	AGCTGGATCCTGCTGAACAAGCAC
NA-R13	141	GCATGTCGACTGGGTAAACCTTGG
NA-R14	207	GCATGTCGACCTCGAGGGAATC
NA-R15	327	GCATGTCGACACCATCTGGGGCTG
NA-R16	468	GCATGTCGACCAGCATTGTTATTAG
NA-R17	624	GCATGTCGACCAGGAGAATTTCCCCTAC
NA-R18	753	GCATGTCGACCAGATTTCTTAGTGACAG
NA-R19	810	GCATGTCGACCGTTGTACTGTTTTGTGG
NA-R20	864	GCATGTCGACCCCCGAAATTTCCTTGG
NA-R22	1047	GCATGTCGACTGTCTTTGAATTGTGGATC
NA-R23	1146	GCATGTCGACAAGGCTGAGCTTCATCAG
NA-R24	1185	GCATGTCGACGAAGAGTCACAGTGGGCTG
NO-F	1	AGCTGGATCCATGTCTGATAATGG
NO-F1	151	AGCTGGATCCGCGTCTTGGTTCACAG
NO-F2	397	AGCTGGATCCTGGGTTGCAACTGAG
NO-F3	661	AGCTGGATCCGCGCTATTGCTGCTAG
NO-F4	949	AGCTGGATCCGGAATGTCACGCATTG
NO-R1	402	AGCTGTCGACAACCCATACGATGC
NO-R2	654	AGCTGTCGACAGTTTCACCACCTCC
NO-R3	933	AGCTGTCGACACTTGGAGCAAATTG
NO-R	1266	AGCTGTCGACTTATGCCTGAGTTG

[a] F, forward primer; R, reverse primer.

[b] Location of the first base at the 5′ end of the primer on the nucleocapsid gene.

[c] The underlined sequences are the restriction site and protection nucleotides.

(Invitrogen). The full-length N gene was amplified from the cDNA templates with the primer pair NO-F and NO-R. The PCR products were purified with Montage PCR reagent sets (Millipore Corporation), digested with *Bam*HI and *Sal*I (Promega Corporation), and ligated into similarly digested plasmid pET-32a (Novagen). The ligation products were transformed in BL21 (DE3; Novagen Inc.), and the recombinant clones were identified by PCR and confirmed by sequencing. Various fragments spanning different putative epitopes (Table 2) were amplified from the recombinant clone with the full-length N protein gene by different primer combinations and cloned into pET-32a as described above. To produce the target proteins, the recombinant clones were cultured at 37 °C to $A_{600} = 0.6$ and then induced with isopropyl β-D-thiogalactoside (Sigma). The proteins were then purified by use of Ni-NTA agarose (Qiagen) according to the manufacturer's instructions.

IMMUNOBLOTTING ANALYSIS OF THE PROTEINS

For immunoblotting analysis, the proteins were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred to nitrocellulose membranes. The membranes were blocked in Tris-buffered saline-Triton (150 mmol/L NaCl; 50 mmol/L bis-Tris, pH 7.5; 0.5 mL/L Triton X-100) containing nonfat milk (10 mL/L) for 2 h at room temperature and then incubated with pooled patient sera (1:200 dilution) for 2 h. After three washes in Tris-buffered saline-Triton, the membranes were incubated with horseradish peroxidase-conjugated goat anti-human IgG (Santa Cruz Biotechnology) for 2 h and then washed with Tris-buffered saline-Triton three more times. The membranes were developed with diaminobenzidine substrate for ~5 min, and the reaction was stopped by the addition of water.

PROTEIN MICROARRAY ANALYSIS

The purified proteins were dissolved at different concentrations (400, 100, 25.5, and 0.5 mg/L) in phosphate-buffered saline (PBS; pH 7.5) containing 400 mL/L glycerol and printed in triplicate on silylated glass slides (CEL Associates, Inc.) by use of the GSI Flexys arrayer (Genomic Solutions). Goat anti-human IgG and human IgG were printed as positive controls, and cell lysates of *E. coli* BL21 (transformed with pET-32a) and PBS (containing 400 mL/L glycerol) were printed as negative controls (see Fig. S1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol50/issue6/). Printed slides were stored at 4 °C and used within 2 weeks after being printed. To eliminate antibodies against *E. coli* in the sera, all of the sera were incubated with lysates of *E. coli* BL21 carrying pET-32a before microarray analysis. For microarray probing, the slides were first blocked with 30 mL/L nonfat milk in PBS for 1 h and then incubated with diluted sera (1:200) for another 1 h at room temperature. The incubated slides were washed three times with PBS containing 1 mL/L Tween-20 and once with PBS and then incubated for another 2 h with the fluorescently labeled secondary antibody (goat anti-human IgG) generated by use of a Cy5 antibody labeling reagent set (Amersham Biosciences). The resulting slides were washed with PBS-Tween twice and PBS once, followed by a 2-min ethanol rinse. After being dried with hot air, the slides were scanned with the GenePix Personal 4100A Scanner (Axon Instruments). Each analysis was repeated at least once.

The scanned images were processed, and the generated data were further analyzed by the GenePix Pro 4.0 software (Axon Instruments) and Microsoft Excel software. The fluorescence signal of each spot was calculated as the median fluorescence intensity (FI) minus the median local background intensity. Spots with bad signals were rejected, and those with signal values <1 were set as 1 to reflect a spot signal close to 0 (see Table S1 in the online Data Supplement). The mean human IgG signal from all of the slides was used to calibrate the signals of other spots. The spot signals for each protein fragment in the replicated hybridizations were then averaged and ready for further analysis. The normalized datasets were logarithmically transformed (base 2) and clustered by CLUSTER (16) and viewed by TREEVIEW (http://genome-www4.stanford.edu/MicroArrayy/SMDyrestech.html). To determine the positive serum number of each protein fragment, the cutoff value of each fragment was calculated as the mean value plus 3 SD of the FI from the negative controls.

For further analysis of the epitope locations and antigenicities of different regions, pooled sera (10 of the 52 tested sera) were mixed with excessive amount of specific fragments to eliminate the corresponding antibodies and then probed with protein microarrays.

VALIDATION OF MICROARRAY ANALYSIS BY ELISA

To further validate the results of the microarray analysis, fragment-specific ELISAs were used. Briefly, 96-well plates were coated overnight with the purified proteins at 4 °C. After the plates were blocked, sera from SARS patients were added for incubation at room temperature for 1 h; the plates were then washed with PBS-Tween. The results were visualized by horseradish peroxidase-conjugated goat anti-human IgG and tetramethylbenzidine.

Results

PRODUCTION AND PURIFICATION OF N PROTEIN AND ITS FRAGMENTS

On the basis of the predicted epitopes, 33 primers were designed and synthesized and used to amplify fragments spanning these epitopes (Table 1). The cDNA for a total of 33 fragments of the N protein was cloned and expressed (Table 2). The positions of these fragments on the N protein are depicted in Fig. 1, and the details of these fragments are listed in Table 2. After optimization of the expression and purification conditions, all of the proteins produced in either soluble form or as inclusion bodies were successfully purified as single bands on sodium dodecyl sulfate–polyacrylamide gel electrophoresis for microarray analysis.

IMMUNOBLOTTING ANALYSIS OF THE PROTEINS

To locate the linear epitopes in N protein and its fragments, the proteins produced were analyzed with pooled sera (15 of the 52). The results (data not shown) indicated that most of the proteins could react with pooled sera. Fragments that did not react with pooled sera were N114, N214, N215, N315, N416, N822, N922, N1022, F14, F15, N12, and N5, indicating that there were no linear epitopes on these fragments. When equal quantities of proteins were analyzed by immunoblotting, fragments N4, N8, N824, N11, and N13 appeared to have higher reactivities, and fragments N417, N517, N9, and N10 appeared to have lower reactivities, as indicated by the amount of gray that developed on the immunoblot.

Table 2. Detailed information for the fragments.

Fragment name	Primer combination[a]	Start position, amino acid number	Stop position, amino acid number	Fragment length, amino acids
F14	NO-F + NA-R14	1	69	69
F15	NO-F + NA-R15	1	109	109
N1	NO-F + NO-R1	1	134	134
F17	NO-F + NA-R17	1	208	208
N2	NO-F + NO-R2	1	218	218
N3	NO-F + NO-R3	1	311	311
N4	NO-F + NO-R	1	422	422
N114	NA-F1 + NA-R14	18	69	52
N214	NA-F2 + NA-R14	31	109	79
N215	NA-F2 + NA-R15	31	134	104
N5	NO-F1 + NO-R1	51	134	84
N6	NO-F1 + NO-R2	51	218	168
N7	NO-F1 + NO-R3	51	311	261
N8	NO-F1 + NO-R	51	422	372
N315	NA-F3 + NA-R15	54	109	56
N416	NA-F4 + NA-R16	72	156	85
N417	NA-F4 + NA-R17	72	208	137
N9	NO-F2 + NO-R2	133	218	86
N10	NO-F2 + NO-R3	133	311	179
N11	NO-F2 + NO-R	133	422	290
N517	NA-F5 + NA-R17	135	208	74
N12	NO-F3 + NO-R3	221	311	91
N13	NO-F3 + NO-R	221	422	202
N822	NA-F8 + NA-R22	249	349	101
N823	NA-F8 + NA-R23	249	382	134
N824	NA-F8 + NA-R24	249	395	147
N922	NA-F9 + NA-R22	273	349	77
N923	NA-F9 + NA-R23	273	382	110
N924	NA-F9 + NA-R24	273	395	123
N1022	NA-F10 + NA-R22	292	349	58
N1023	NA-F10 + NA-R23	292	382	91
N1024	NA-F10 + NA-R24	292	395	104
N14	NO-F4 + NO-R	317	422	106

[a] Primer combinations used to amplify the corresponding RNA.

CORRELATION BETWEEN CAPTURED IgG AND FI

To explore the correlation between the FI and the IgG quantities captured by the printed antigens, human IgG in increasing concentrations was printed on slides to act as an internal standard for calibration. As the results showed, there was a good correlation ($r^2 = 0.9987$) between FI and IgG concentration (Fig. 2A), indicating that the FI can reflect the amount of IgG captured by the printed antigens. Because each pin was estimated to transfer a 40-pL sample, the calibration curve of FI to the amount of IgG could be calculated (Fig. 2B). This curve could be used to estimate the amounts of antibodies captured by antigens on the microarray. In general, fragments with higher reactivities captured more antibodies. For example, fragment N4 captured ~9.65 pg of IgG, three times the amount captured by F17 (see Fig. S2 in the online Data Supplement). The correlation between the FI and printed antigen concentration was lower than that of the printed human IgG (data not shown). However, the
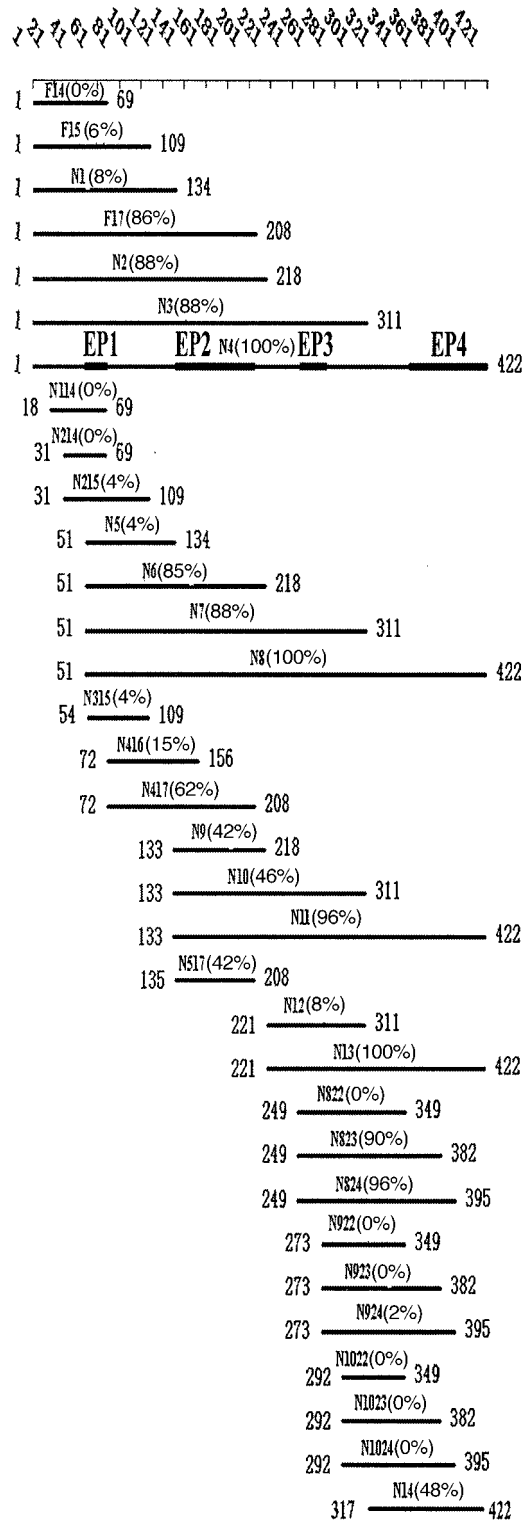

Fig. 1. Positions of the fragments on the SARS-CoV N protein.

The *line* with *scales* at the *top* represents the full-length N protein (*N4*), with the *numbers* indicating amino acid positions. The *other lines* represent the fragments, and the *numbers* on both *ends* of each fragment indicate the starting and ending amino acid positions. The fragment name is *above* the *line*; the *numbers* in *parentheses* indicate the percentages of sera positive for each fragment by protein microarray analysis. The four important epitope regions (*EP1–EP4*) are located at amino acids 51–71, 134–208, 249–273, and 349–422, respectively, as indicated by the *thick lines*.
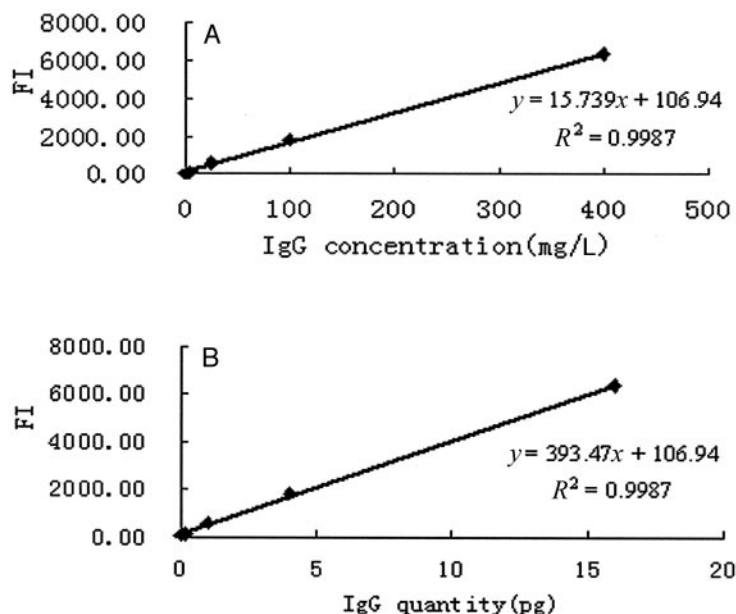
Fig. 2. Correlation between FI and captured IgG.

(*A*), correlation between FI and concentration of printed IgG. Human IgG was printed in increasing concentrations and probed with Cy5-labeled goat anti-human IgG. The FI was plotted as a function of the concentration of printed human IgG to generate a dose–response curve, and a good correlation coefficient (close to 1) was obtained. (*B*), correlation between FI and amount of printed human IgG.

In the figure:

Panel A: $y = 15.739x + 106.94$, $R^2 = 0.9987$, x-axis: IgG concentration (mg/L)

Panel B: $y = 393.47x + 106.94$, $R^2 = 0.9987$, x-axis: IgG quantity (pg)

correlation was good for certain fragments that could react with most (>80%) of the tested sera (e.g., $r^2 = 0.9823$ for N4).

VALIDATION OF PROTEIN MICROARRAY ANALYSIS

Fragments with different reactivities in microarray assays were selected and analyzed by fragment-specific ELISAs. The absorbance values in ELISA were used for regression analysis against the FI in microarray analysis. The results showed that there was good correlation between IgG absorbance and FI, particularly for fragments that could react with most (>80%) of the sera (data not shown). Fragments with poor reactivity in microarray analysis had a lower correlation coefficient (e.g., $r^2 = 0.8583$ for N1), whereas for those that could react with 100% (52 of 52) of the sera, the correlation was high (e.g., $r^2 = 0.9993$ for N4).

PROFILE OF THE ANTIGENICITY OF N PROTEIN FRAGMENTS

The normalized fluorescence values were logarithmically transformed (base 2) and clustered by CLUSTER and graphed by TREEVIEW (Fig. 3). Most of the fragments were clustered into four main groups, among which the fragments in group A had the highest reactivity (>80%), fragments in group B reacted with 40–60% of the tested sera, fragments in group C had lower reactivity (0–40%), and group D included only F17, which also reacted with most (86%) of the sera. N13 (amino acids 221–422) and N2 (1–218) were approximately the same length but reacted with 100% and 86% of the sera, respectively (Fig. 1), inferring that there were important antigen determinants at both termini of the N protein but that the antigenicity at the COOH terminus was higher than that at the NH_2 terminus. In addition, as predicted by DNASTAR software, there were more putative epitopes located at the COOH terminus of the N protein *(17)*.

REGIONS SPANNING AMINO ACIDS 1–50 AND 218–311 ARE NOT IMPORTANT FOR THE EXPOSURE OF EPITOPES

Fragments N7 (amino acids 51–311) and N3 (1–311) reacted with the same percentage of serum percentile (88%), indicating that the 50 amino acids at the NH_2 terminus of N3 were not important for the exposure of epitopes. The results for N6 (amino acids 51–218) and N8 (51–422), which were formed by the deletion of these 50 amino acids from N2 (amino acids 1–218) and N4 (1–422), respectively (Fig. 1), also supported this conclusion. Despite a longer peptide, N3 reacted with the same number of sera as N2, suggesting that the region spanning amino acids 218–311 was not important for antigenicity.

REGIONS SPANNING AMINO ACIDS 50–70 AND 138–208 HAVE IMPORTANT EPITOPES

Fragment F17 (amino acids 1–208) had positive reactions with 86% (45 of 52) of the tested sera, but fragment N1 (1–134) reacted with only 8% (4 of 52) of the tested sera, indicating that the region spanning amino acids 135–208 contains important antigenic determinants. Fragment N417 (amino acids 72–208) reacted with 62% (32 if 52) of the sera, which was lower than the reactivity of fragment F17, demonstrating that the first 70 amino acids might include epitopes that increase the reactivity of F17. Because the first 50 amino acids (amino acids 1–50) were not important for antigenicity, the epitopes were likely located at the region of amino acids 50–70.

REGIONS SPANNING AMINO ACIDS 249–273 AND 349–422 CONTAIN IMPORTANT EPITOPES

Fragments N822 (amino acids 249–349), N823 (249–382), and N824 (249–395) reacted with 0 (0%), 47 (90%), and 50 (96%) of the 52 sera, inferring that there are antigen determinants in the region of amino acids 349–395. Fragment N923 (amino acids 273–382), which was formed by
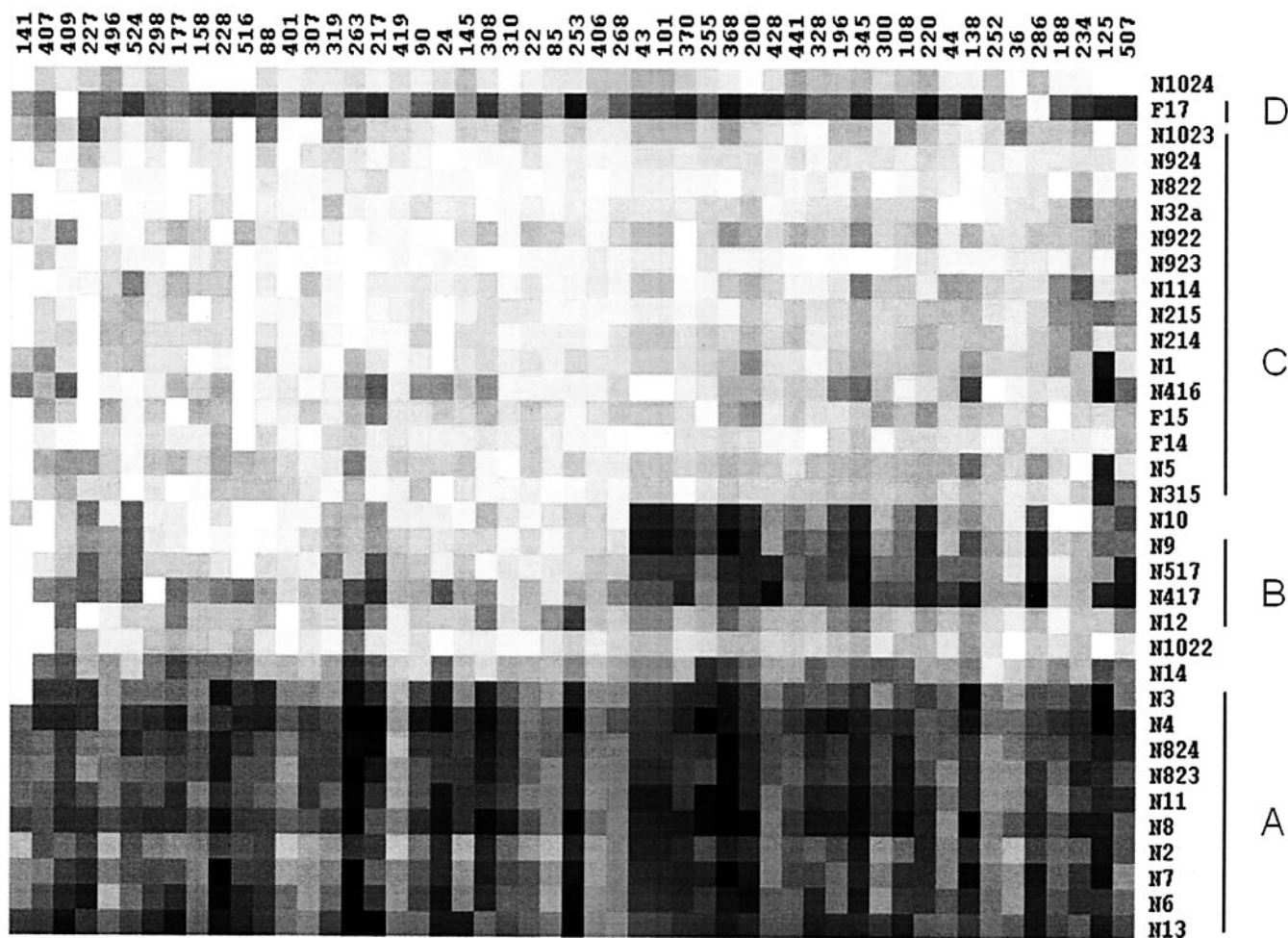
Fig. 3. Reactivities of all of fragments printed at the concentration of 400 mg/L.

The fluorescence values were normalized and logarithmically transformed (base 2); the cutoff value 8 was subtracted from each result. The values were then clustered by CLUSTER software and viewed with TREEVIEW software. *Black squares* represent positive values, *white squares* indicate negative values, and *gray squares* indicate intermediate degrees of FI.

deletion of 25 amino acids (249–273) from the NH_2 terminus of N823, did not react with any of the tested sera, suggesting that this 25-amino acid region is very important. N14 (amino acids 317–422), which was formed by deletion of 26 amino acids at the NH_2 terminus and an elongation of 28 amino acids (396–422) at the COOH terminus of N1024, reacted with 25 (48%) of the sera, which further confirmed the important role of this 28-amino acid region in the antigenicity of N14.

FOUR IMPORTANT REGIONS IDENTIFIED AND THEIR RELATIONSHIPS

From the analysis above, four regions containing important epitopes were identified in the N protein in this study: EP1 (amino acids 51–71), EP2 (134–208), EP3 (249–273), and EP4 (349–422). Fragments that contained only EP1 had lower or no reactivities; for example, F14 and N1 reacted with only 0% and 8% of the tested sera, respectively (Fig. 1). Fragments that contained only EP2 (i.e., N9 and N517) reacted with ~40% of the sera. Fragment N2,

which spanned both EP1 and EP2, could react with most of the sera (46 of 52; 88%), which indicated a cooperative relationship between EP1 and EP2. Fragment N9 (amino acids 133–218) was reactive in both the immunoblotting and microarray analyses, indicating that EP2 contains both conformational and linear epitopes.

Fragments that contained only EP4 reacted with approximately one-half of the tested sera, and those that contained only EP3 had no reactivity. Fragments that spanned both of them, however, reacted with 100% (52 of 52) of the tested sera, suggesting that there might be some relationship between EP3 and EP4 for the formation of conformational epitopes. Fragments N922, N923, and N924, which contained only EP4, had no reactivity in protein chip analysis; however, N923 and N924 could react with pooled sera in immunoblotting analysis, indicating that EP4 contains only linear epitopes. Fragments N822, N12, and N922 had no reactivity in both analyses, implying that EP3 could not form epitopes alone.

NOT ALL EPITOPES IDENTIFIED IN THE N PROTEIN ARE EXPOSED IN THE FULL-LENGTH PROTEIN

To further analyze the antibodies against different regions of the N protein, we pooled sera, mixed them with excessive amounts of specific fragments, and then probed the sera with protein microarrays. When mixed with the full-length protein, the pooled sera still reacted with fragments N13 and N824, indicating that in the full-length protein, not all of the identified epitopes are exposed (Table 3). Fragment N13 reacted with the serum to which N824 had been added and vice versa; we thus could infer that they exposed different epitopes.

Discussion

In recent years, protein microarray technology has become a particularly powerful tool for high-throughput gene function analysis (18, 19). Although there are still challenges in protein microarray technology, major advances have been achieved in this field (20–22). In the present study, fragments of the N protein of the SARS-CoV were purified and printed on glass slides and probed with patient sera to analyze their reactivity. The proteins printed on the slides retained their conformations and were recognized by the relevant antibodies in complex solutions.

In the present study, we analyzed the antigenicity of different regions of the SARS-CoV N protein and identified four important regions (EP1 to EP4). Wang et al. (15) analyzed epitopes of the N protein, using synthetic pep-

tides, and found that N66 and N371–N404 were important epitopes. Notably, peptides N371 and N385, located at the COOH terminus of the N protein, inhibited the binding of antibodies to the SARS-CoV lysate and bound to antibodies in >94% of the tested sera. N385 had the highest affinity for forming peptide–antibody complexes with SARS serum. In the present study, EP2 (amino acids 51–71), which overlaps with N66, was found to be important for antigenicity. In our microarray analyses, the regions located at the COOH terminus were found to be more important than those at the NH$_2$ terminus, which further confirmed the results of a previous study. Wang et al. (17) found that epitopes located at amino acids 161–182 and 471–390 were highly immunogenic, which was also consistent with our results. As indicated in this report, EP4 (amino acids 349–422) carried linear epitopes that gave negative results in microarray analyses but were positive in immunoblotting. A total of four important regions were identified in the N protein, and fragments containing only one of these regions had no or low reactivity, whereas those spanning two or more of these regions had high reactivity. Thus, structural requirements seem to be important for antigenicity of the N protein because a cooperative action is required between these regions. This also showed the important role of conformational epitopes in the antigenicity of the N protein, which could not be obtained by synthetic peptides.

Shi et al. (14) used a N-protein-based antigen-capturing ELISA to analyze antibodies to this protein. They found that anti-N-protein antibodies could be detected in 68.4% of probable SARS patients 6–10 days after illness and in 89.6% of the patients 11–61 days after illness, indicating the high immunogenicity of the N protein and early appearance of its antibodies. The antibodies against the N protein were produced in all SARS patients and appeared in early stages of this disease, which implies the possible use of this protein for early and accurate diagnosis of this disease.

There are three important domains in the N protein, the second of which is a RNA-binding domain (23, 24). In the genome of the murine hepatitis virus, the RNA-binding domain is located at amino acids 175–231 (25). Sequence analysis indicates that the RNA-binding domain of the SARS-CoV N protein is at amino acids 178–205 (26). Wang et al. (15) found that N177 (amino acids 177–198) was highly immunogenic, and in the present study, EP2 overlaps this region, indicating that this region is important both in its RNA-binding function and its immunogenicity. Motif scanning of the SARS-CoV predicted a bipartite nuclear localization signal located at amino acids 373–390, suggesting that the N protein may play some special role in the pathogenicity of this new CoV (27). In this region, important epitopes (N371 and N385) were identified (15). These predicted functional sites overlap important epitopes, implying the possibility of using of this protein as a vaccine candidate.

Table 3. Results for pooled sera mixed with the specific fragments.

Fragment[a]	No fragment added	Fragments mixed with the sera before reaction[b]							
		N11	N13	N2	N4	N3	N6	N8	N824
F17	+	−	+	−	−	−	−	−	+
N417	+	−	−	−	−	−	−	−	+
N517	+	−	+	−	−	−	−	−	+
N823	+	+	−	+	+	+	+	−	−
N824	+	+	−	+	+	+	+	+	−
N2	+	−	+	−	−	−	−	−	+
N3	+	−	+	−	−	−	−	−	+
N4	+	+	+	+	−	+	+	−	+
N6	+	−	+	−	−	−	−	−	+
N7	+	−	+	−	−	−	−	−	+
N8	+	+	+	+	−	+	+	−	+
N9	+	−	+	−	−	−	−	−	+
N10	+	−	+	−	−	−	−	−	+
N11	+	−	+	+	−	+	+	−	+
N12	+	−	−	−	−	−	−	−	−
N13	+	+	−	+	+	+	+	+	+

[a] Fragments that can react with pooled sera before the sera were mixed with excessive amounts of specific fragments.

[b] Excessive amounts of specific fragments were mixed with the pooled sera. The sera were then probed with microarray. +, positive reaction; −, negative reaction.

In conclusion, in the present study, protein microarray technology was used to analyze the antigenicity of the N protein for the first time. By use of comprehensive microarray analyses, we identified four important regions of the SARS-CoV N protein and confirmed that the COOH terminus has higher immunogenicity than the NH_2 terminus. We also demonstrated that the N protein of SARS-CoV is highly immunogenic and could be used as diagnostic antigen. The fragments that reacted to all of the tested sera in this study are being used as antigens to immunize animals to prepare antibodies to investigate their possible use as vaccine candidates.

References

1. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003;348:1967–76.
2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003;348:1953–66.
3. Falsey AR, Walsh EE. Novel coronavirus and severe acute respiratory syndrome. Lancet 2003;361:1312–3.
4. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, et al. The genome sequence of the SARS-associated coronavirus. Science 2003;300:1399–404.
5. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003;300:1394–9.
6. Enserink M. SARS in China. China's missed chance. Science 2003;301:294–6.
7. Narayanan K, Makino S. Characterization of nucleocapsid-M protein interaction in murine coronavirus. Adv Exp Med Biol 2001;494:577–82.
8. Narayanan K, Maeda A, Maeda J, Makino S. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J Virol 2000;74:8127–34.
9. Lai MM. Coronavirus: organization, replication and expression of genome. Annu Rev Microbiol 1990;44:303–33.
10. Xu HY, Lim KP, Shen S, Liu DX. Further identification and characterization of novel intermediate and mature cleavage products released from the ORF 1b region of the avian coronavirus infectious bronchitis virus 1a/1b polyprotein. Virology 2001;288:212–22.
11. Tahara SM, Dietlin TA, Bergmann CC, Nelson GW, Kyuwa S, Anthony RP, et al. Coronavirus translational regulation: leader affects mRNA efficiency. Virology 1994;202:621–30.
12. Nelson GW, Stohlman SA, Tahara SM. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. J Gen Virol 2000;81:181–8.
13. Seo SH, Wang L, Smith R, Collisson EW. The carboxyl-terminal 120-residue polypeptide of infectious bronchitis virus nucleocapsid induces cytotoxic T lymphocytes and protects chickens from acute infection. J Virol 1997;71:7889–94.
14. Shi Y, Yi Y, Li P, Kuang T, Li L, Dong M, et al. Diagnosis of severe acute respiratory syndrome (SARS) by detection of SARS coronavirus nucleocapsid antibodies in an antigen-capturing enzyme-linked immunosorbent assay. J Clin Microbiol 2003;41:5781–2.
15. Wang J, Wen J, Li J, Yin J, Zhu Q, Wang H, et al. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. Clin Chem 2003;49:1989–96.
16. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863–8.
17. Wang J, Ji J, Ye J, Zhao X, Wen J, Li W, et al. The structure analysis and antigenicity study of the N protein of SARS-CoV. Genomics Proteomics Bioinformatics 2003;1:145–54.
18. de Wildt RM, Mundy CR, Gorick BD, Tomlinson IM. Antibody arrays for high-throughput screening of antibody-antigen interactions. Nat Biotechnol 2000;18:989–94.
19. Knezevic V, Leethanakul C, Bichsel VE, Worth JM, Prabhu VV, Gutkind JS, et al. Proteomic profiling of the cancer microenvironment by antibody arrays. Proteomics 2001;1:1271–8.
20. Wilson DS, Nock S. Recent developments in protein microarray technology. Angew Chem Int Ed Engl 2003;42:494–500.
21. Haab BB. Advances in protein microarray technology for protein expression and interaction profiling. Curr Opin Drug Discov Dev 2001;4:116–23.
22. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, et al. Global analysis of protein activities using proteome chips. Science 2001;293:2101–5.
23. Parker MM, Masters PS. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. Virology 1990;179:463–8.
24. Cologna R, Spagnolo JF, Hogue BG. Identification of nucleocapsid binding sites within coronavirus-defective genomes. Virology 2000;277:235–49.
25. Kuo L, Masters PS. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J Virol 2002;76:4987–99.
26. Shi L, Rui W, Lee M. [Preliminary analysis of the structure and function of SARS putative nucleocapsid protein]. http://cmbi.bjmu.edu.cn/cmbidata/sars/html/41.pdf (accessed June 2003).
27. Shi L, Rui W, Lee M. [Nuclear targeting sequence in SARS nucleocapsid protein]. http://cmbi.bjmu.edu.cn/cmbidata/sars/html/10.pdf (accessed June 2003).