

# Elucidation of the stability and functional regions of the human coronavirus OC43 nucleocapsid protein

Chun-Yu Huang,<sup>1,2,3</sup> Yen-Lan Hsu,<sup>4</sup> Wan-Ling Chiang,<sup>5</sup> and Ming-Hon Hou<sup>1,2,3\*</sup>

<sup>1</sup>Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

<sup>2</sup>Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 402, Taiwan

<sup>3</sup>Department of Life Science, National Chung Hsing University, Taichung 402, Taiwan

<sup>4</sup>Institute of Biomedical Science, Academia Sinica, Taipei 115, Taiwan

<sup>5</sup>Institute of Biochemistry, National Chung Hsing University, Taichung 402, Taiwan

Received 2 July 2009; Revised 6 August 2009; Accepted 10 August 2009

DOI: 10.1002/pro.225

Published online 18 August 2009 proteinscience.org

**Abstract:** Human coronavirus OC43 (HCoV-OC43) is one of the causes of the “common cold” in human during seasons of cold weather. The primary function of the HCoV-OC43 nucleocapsid protein (N protein) is to recognize viral genomic RNA, which leads to ribonucleocapsid formation. Here, we characterized the stability and identified the functional regions of the recombinant HCoV-OC43 N protein. Circular dichroism and fluorescence measurements revealed that the HCoV-OC43 N protein is more highly ordered and stabler than the SARS-CoV N protein previously studied. Surface plasmon resonance (SPR) experiments showed that the affinity of HCoV-OC43 N protein for RNA was approximately fivefold higher than that of N protein for DNA. Moreover, we found that the HCoV-OC43 N protein contains three RNA-binding regions in its N-terminal region (residues 1–173) and central-linker region (residues 174–232 and 233–300). The binding affinities of the truncated N proteins and RNA follow the order: residues 1–173–residues 233–300 > residues 174–232. SPR experiments demonstrated that the C-terminal region (residues 301–448) of HCoV-OC43 N protein lacks RNA-binding activity, while crosslinking and gel filtration analyses revealed that the C-terminal region is mainly involved in the oligomerization of the HCoV-OC43 N protein. This study may benefit the understanding of the mechanism of HCoV-OC43 nucleocapsid formation.

**Keywords:** human coronavirus; OC43 strain; nucleocapsid protein; Ribonucleocapsid; stability; RNA-binding regions; oligomerization domain

---

Additional Supporting Information may be found in the online version of this article.

**Abbreviations:** CoV, coronavirus; E, envelope; HCoV, human coronavirus; IBV, avian infectious bronchitis virus; M, membrane; MHV, murine hepatitis virus; N protein, nucleocapsid protein; RNP, ribonucleoprotein; S, spike; SARS, severe acute respiratory syndrome; SR-rich, serine and arginine-rich;  $T_m$ , melting temperature.

The Symposium of the Protein Society and Experimental Biology Annual Meeting.

Grant sponsor: NCS; Grant number: 97-2311-B-005-003-MY3.

\*Correspondence to: Ming-Hon Hou, Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung 402, Taiwan. E-mail: mhho@dragon.nchu.edu.tw

## Introduction

The strain OC43 of the type II coronavirus family (HCoV-OC43) was first identified in the 1960s and is responsible for ~20% of all “common colds” in humans.<sup>1,2</sup> Although HCoV-OC43 infections are generally mild, more severe upper and lower respiratory tract infections like bronchiolitis and pneumonia have been documented, especially in infants, elderly individuals, and immunocompromised patients.<sup>1,3,4</sup> Moreover, there have been reports that clusters of HCoV-OC43 infections cause pneumonia in otherwise healthy adults.<sup>2,5</sup> Several studies have reported that both neurotropism and neuroinvasion of HCoV are associated with multiple sclerosis, especially in the case of the OC43 strain.<sup>6</sup>

CoV particles have an irregular shape, which is determined by an outer envelope with distinctive, “club-shaped.” These peplomers give the virus a crown-like (corona) appearance.<sup>7</sup> The viral genome of coronaviruses consists of positive sense, single-stranded RNA of ~30 kb. The genome contains several genes that encode structural and nonstructural proteins that are required for progeny virion production.<sup>1</sup> The virion envelope surrounding the nucleocapsid contains the following structural proteins: S (spike) protein, M (membrane), and E (envelope). A third glycoprotein, HE (hemagglutinin-esterase), is present in most class II coronaviruses.<sup>8,9</sup> A helical nucleocapsid exists in the center of the viral particle.<sup>10–12</sup> Nucleocapsid protein, the major structural protein of CoVs, binds to the viral RNA genome to form the virion core, which leads to the formation of a ribonucleoprotein (RNP) complex or to a long helical nucleocapsid structure.<sup>13,14</sup> The formation of the RNP is important for maintaining the RNA in an ordered conformation suitable for replication and transcription of the viral genome.<sup>13,15–17</sup> Previous studies have shown that the CoV N protein is involved in the regulation of cellular processes such as gene transcription, actin reorganization, host cell cycle progression, and apoptosis.<sup>18–21</sup> It has also been shown to act as an RNA chaperone.<sup>22</sup> Moreover, the N protein is an important diagnostic marker and the most “immuno-dominant” antigen in a host immune response.<sup>17,23,24</sup>

Previous studies have revealed that the N- and C-terminal domains of the CoV N proteins, including those of SARS-CoV, murine hepatitis virus (MHV), and avian infectious bronchitis virus (IBV), are responsible for RNA binding and oligomerization, respectively.<sup>25–30</sup> The central region of the N protein has also been shown to contain an RNA-binding region and the primary sites of phosphorylation.<sup>28,31,32</sup> Phosphorylation of the N protein has been shown to play an important role in virus biology.<sup>33,34</sup> To clarify the molecular mechanism of ribonucleocapsid formation by CoVs, structures of truncated fragments of N protein including N-terminal and C-terminal domains have been published.<sup>26,35–37</sup> Despite the conservation of some motifs, CoV N proteins from different strains often show quite different properties primarily due to their low sequence homology.<sup>30</sup> For example, Saikatendu *et al.* found that the structures of SARS-CoV and IBV N-terminal domains shared several common features, in addition to many subtle structural differences.<sup>36</sup> On the basis of the different crystal packing of these two structures, they suggested that the two viruses most likely use different modes of oligomeric self-association during RNP core formation.

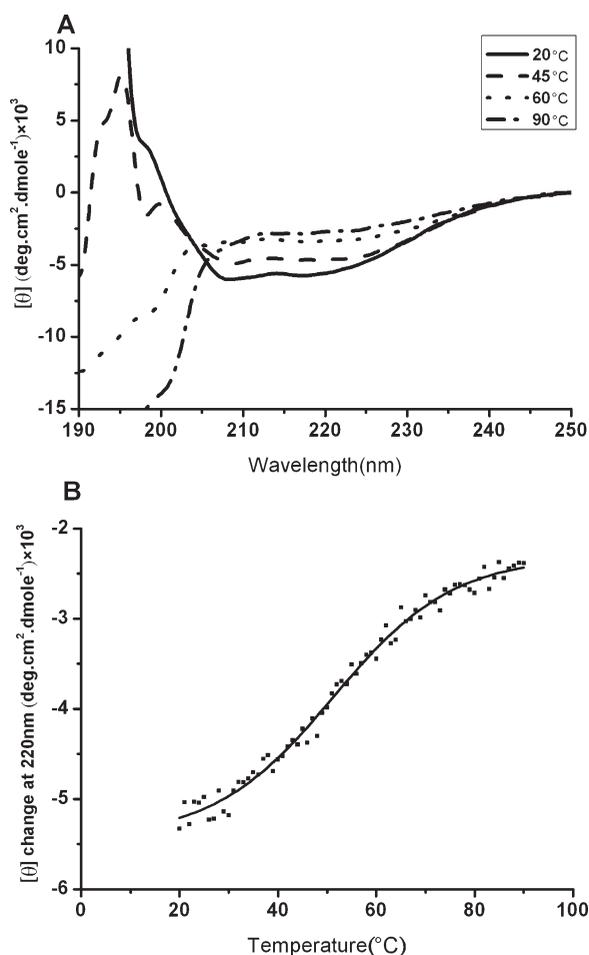
The N protein of HCoV-OC43, which has a molecular weight of 50 kDa, is highly basic (pI, 10), and it shows strong hydrophilicity.<sup>38</sup> The N protein of HCoV-OC43 shows only 26–30% amino acids similarity to

the N protein of other CoV strains.<sup>30</sup> Until now, few studies have examined the biochemical properties of the HCoV-OC43 N protein. In this study, we characterized the stability and biochemical properties of the recombinant full-length N protein from HCoV-OC43. RNA-binding and oligomerization regions of the HCoV-OC43 N protein were identified.

## Results

### Stability studies of recombinant HCoV-OC43 nucleocapsid protein

Conformational changes of the HCoV-OC43 N protein in response to temperatures were monitored using CD spectroscopy. As shown in Figure 1(A), the CD spectra of HCoV-OC43 N protein were scanned from 190 to 250 nm at 20, 45, 60, and 90°C. Analysis by the SELECON program revealed that HCoV-OC43 N



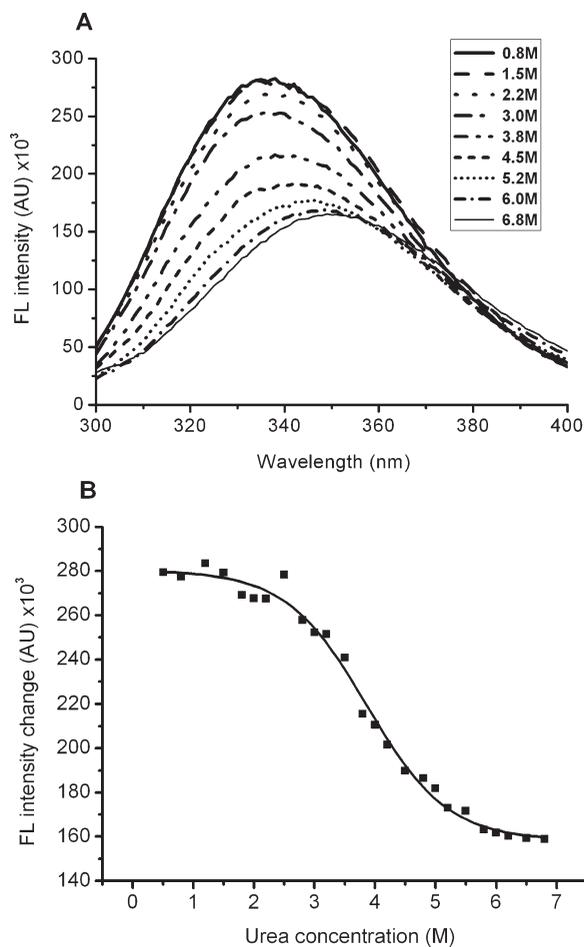
**Figure 1.** (A) CD spectra of HCoV-OC43 N protein at various temperatures including 20, 45, 60, and 90°C. The protein concentration was 5  $\mu$ M and the buffer consisted of 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS. (B) Thermostability measurements of HCoV-OC43 N protein monitored by CD spectra. The protein concentration was 3.8  $\mu$ M and the buffer consisted of 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS.

protein contains ~35.1%  $\alpha$ -helices, 11.6%  $\beta$ -sheets, 23.8% turns, and 29.5% random coils at room temperature. At high temperatures, the CD spectra of HCoV-OC43 N protein showed decreased intensity between 210 and 220 nm and a large negative peak at 190 nm, which are characteristic of disordered protein. Previous CD spectra of the full-length N proteins from SARS and 229E CoVs indicated that the N protein is relatively disordered and composed of ~50% nonsecondary structures.<sup>17</sup> Our present results suggest that the conformations of HCoV-OC43 N proteins are more ordered than those of HCoV-229E and SARS-CoV N proteins [Fig. 1(A)]. We also measured the melting temperature ( $T_m$ ) of HCoV-OC43 N proteins using CD [Fig. 1(B)]. Heat denaturation analysis showed that the  $T_m$  of HCoV-OC43 N protein is 52°C under our conditions. The thermal denaturation of HCoV-OC43 N protein is reversible (data not shown).

Denaturant reagents including urea were used to measure the denaturation of HCoV-OC43 N protein. We used tryptophan (Trp) fluorescence to monitor unfolding because the HCoV-OC43 N protein contains five Trp residues [Fig. 2(A)]. The fluorescence emission spectra showed the maximal emission wavelength of the HCoV-OC43 N protein at ~340 nm. Urea-induced unfolding caused a significant decrease in FL intensity at 340 nm and a red shift in the maximal emission wavelength. The unfolding of HCoV-OC43 N protein in the presence of urea was observed by monitoring fluorescence spectra at 340 nm, and the denaturation curve was fitted to a two-state model [Fig. 2(B)].<sup>39</sup> The denaturation and renaturation curves were completely superimposable (data not shown). The denaturant molarity at the midpoint of the transition (called  $C_m$ ) was calculated as 4 M and the unfolding free energy change ( $\Delta G_{N-U}^0$ ) was determined from the linear extrapolation method to be 14.5 kJ/mol (Supporting Information Figure S1).<sup>39</sup>

### Studies of nucleic acid binding by HCoV-OC43 nucleocapsid proteins

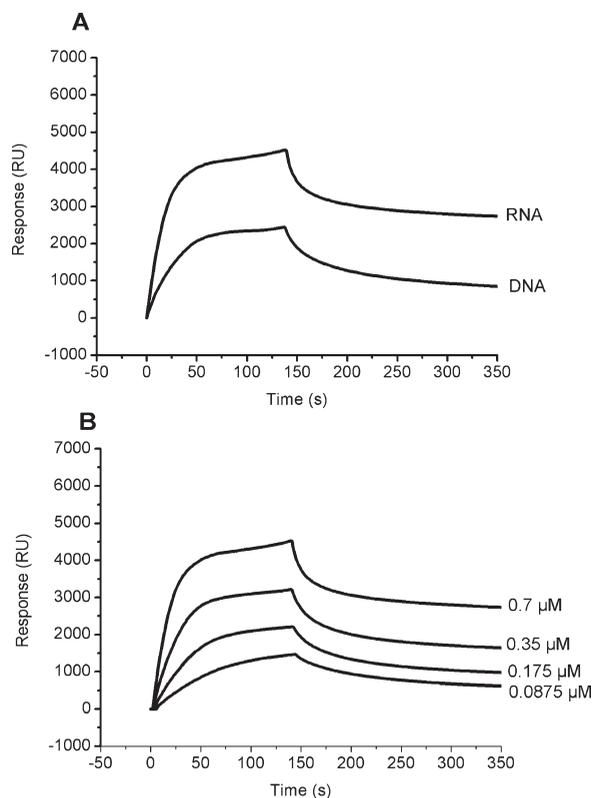
To measure the binding affinity between the HCoV-OC43 nucleocapsid protein and single-stranded nucleic acids, HCoV-OC43 nucleocapsid protein was allowed to interact with biotin-labeled single-stranded RNA or DNA, and the results were monitored by SPR. Depending on the virus strain, there are two to four UCUAA pentanucleotide repeats, with the last repeat being UCUAAAC and termed the intergenic (IG) sequence at the 3' end of the leader.<sup>40,41</sup> Previous studies showed that HCoV N protein has high affinity for the intergenic sequence.<sup>42,43</sup> Therefore, the repeated intergenic sequence of HCoV-OC43, 5'-(UCUAAAC)<sub>4</sub>-3', was used as a probe in our SPR and fluorescence experiments. For DNA, uracil was replaced by thymine. The results show that HCoV-OC43 N protein was able to interact with both DNA and RNA. However, HCoV-OC43 N protein showed a



**Figure 2.** (A) Fluorescence spectra of HCoV-OC43 N protein in Tris-HCl supplemented with different concentrations of urea. The protein concentration was 1  $\mu$ M and the buffer consisted of 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS. (B) Urea-induced unfolding of HCoV-OC43 N protein monitored by fluorescence emission at 340 nm. The protein concentration was 1  $\mu$ M and the buffer consisted of 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS.

higher binding capacity for RNA (~4989 RU) than for DNA (~2803 RU) at the same protein concentration [Fig. 3(A)].

Kinetic experiments were carried out by measuring the binding affinity between HCoV-OC43 N protein and single-stranded RNA or DNA. Four concentrations of N protein were used to determine both the  $k_a$  and  $k_d$ . Figure 3(B) shows the Biacore SPR traces of N protein bound to RNA. The  $k_a$  and  $k_d$  were calculated from the association and dissociation phases of the SPR traces, respectively. The  $k_a$  and  $k_d$  of RNA were  $4.82 \times 10^4 M^{-1} s^{-1}$  and  $5.75 \times 10^{-4} s^{-1}$ , respectively. The kinetic parameters of DNA showed a lower  $k_a$  and a higher  $k_d$  of  $\sim 2.85 \times 10^4 M^{-1} s^{-1}$  and  $1.69 \times 10^{-3} s^{-1}$ , respectively. The dissociation constants ( $K_d$ ) of HCoV-OC43 N protein with target nucleic acid strands were calculated as  $k_d/k_a$  (in M). The  $K_d$  of the N protein bound to RNA was 11.9 nM, which is close



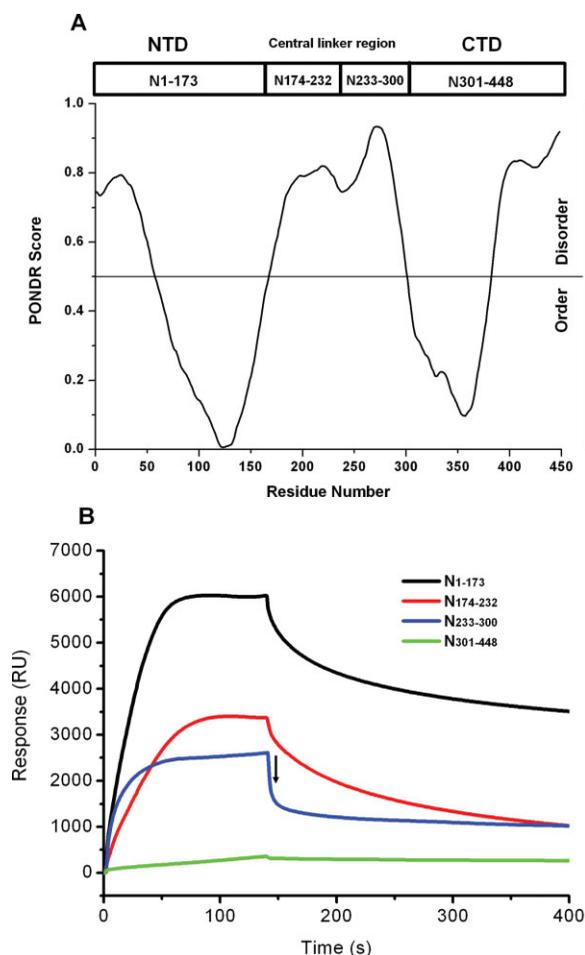
**Figure 3.** (A) Binding capacity sensorgrams of HCoV-OC43 N protein (0.7  $\mu\text{M}$ ) bound to single-stranded RNA or single-stranded DNA. The RNA and DNA sequences immobilized on the SA chip were 5'-(UCUAAAC)4-3' and 5'-(TCTAAAC)4-3', respectively. (B) Sensorgrams of HCoV-OC43 N protein bound to single-stranded RNA. HCoV-OC43 N protein was exposed to RNA at concentration of 0.7, 0.35, 0.175, and 0.0875  $\mu\text{M}$ . The buffer consisted of 50 mM Tris-HCl (pH 7.3) containing 150 mM NaCl and 0.1% CHAPS. The association rate increased with an increasing concentration of target N protein and the dissociation rate was independent of concentration. The fitted data was shown in Supporting Information (Figure S3).

to the previously reported  $K_d$  of  $\sim 14$  nM for the interaction between MHV N protein and RNA.<sup>42</sup> As expected, the  $K_d$  for N protein and DNA was approximately fivefold higher than that for N protein and RNA.

#### Identification of the RNA-binding region of HCoV-OC43 nucleocapsid protein

According to the PONDR prediction, HCoV-OC43 N protein contains two structural domains located at the N-terminal and C-terminal regions, and between them lies a central-linker region that is predicted to be disordered [Fig 4(A)].<sup>30</sup> To further isolate the RNA-binding region of HCoV-OC43 N protein, we expressed and purified the truncated forms of the N protein: one construct contained only the N-terminal domain (N1–173), another contained only the C-terminal domain (N301–448), and two contained different parts of the central-linker region (N174–232 and N233–300).

These domain boundaries are based on the PONDR predictions [Fig. 4(A)]. The binding of each truncated N protein to RNA was analyzed by SPR [Fig. 4(B)]. These studies showed that the N-terminal domain (N1–173) and the central-linker region (N174–232 and N233–300), but not the C-terminal domain, bound to RNA [Fig. 4(B)]. The kinetic constants for association ( $k_a$  in  $M^{-1} s^{-1}$ ) and dissociation ( $k_d$  in  $s^{-1}$ ) of the truncated N proteins with respect to single-stranded RNA have been calculated and listed in Table I. The  $k_a$  values were essentially the same:  $2.18 \times 10^4$  and  $2.48 \times 10^4 M^{-1} s^{-1}$  for N1–173 and N233–300, respectively, and  $7.97 \times 10^3 M^{-1} s^{-1}$  for N174–232. Moreover, the  $k_d$  values of N1–173 and N233–300 were essentially



**Figure 4.** (A) Order/disorder prediction for HCoV-OC43 N protein using the VL3-BA predictor of the PONDR program. (B) Sensorgrams of the truncated forms of HCoV-OC43 N protein including N1–173, N174–232, N233–300, and N301–448 bound to single-stranded RNA. Protein was tested at concentrations of 2  $\mu\text{M}$  in buffer containing 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS. The RNA immobilized on the SA chip was 5'-(UCUAAAC)4-3'. In N233–300, bulk effects caused by the differences in the refractive index of the running buffer and sample solution caused a sharp descent in the curves from 140 s to 142 s; hence, the dissociation measurements were started at 142 s (marked by the arrow).

**Table I.** The Kinetic Parameters of the Binding of Full-Length and Truncated N Proteins to Single-Stranded RNA, 5'-(UCUAAAC)4-3'

Construct	$k_a$ ( $M^{-1} s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$^aK_d$ ( $M$ )
Full-length	$4.82 \pm 0.21 \times 10^4$	$5.75 \pm 0.23 \times 10^{-4}$	$11.9 \pm 1.0 \times 10^{-9}$
N1-173	$2.18 \pm 0.34 \times 10^4$	$1.5 \pm 0.21 \times 10^{-3}$	$68.9 \pm 8.4 \times 10^{-9}$
N174-232	$7.97 \pm 0.62 \times 10^3$	$3.53 \pm 0.32 \times 10^{-3}$	$44.4 \pm 3.7 \times 10^{-8}$
N233-300	$2.48 \pm 0.28 \times 10^4$	$1.58 \pm 0.24 \times 10^{-3}$	$64.1 \pm 7.2 \times 10^{-9}$
N301-448	<sup>b</sup> —	—	—

<sup>a</sup>  $K_d$  values were obtained by dividing  $k_d$  by  $k_a$ .

<sup>b</sup> No significant binding was observed.

the same:  $1.50 \times 10^{-3}$  and  $1.58 \times 10^{-3} s^{-1}$ , respectively. These constants were smaller than that of N174-232 ( $3.53 \times 10^{-3} s^{-1}$ ). The dissociation constants ( $K_d$ ) of all truncated N proteins with RNA were calculated as  $k_d/k_a$  (in  $M$ ) and they follow the order: N174-232 > N1-173-N233-300. Our results suggest that the N protein of HCoV-OC43 contains three RNA-binding regions.

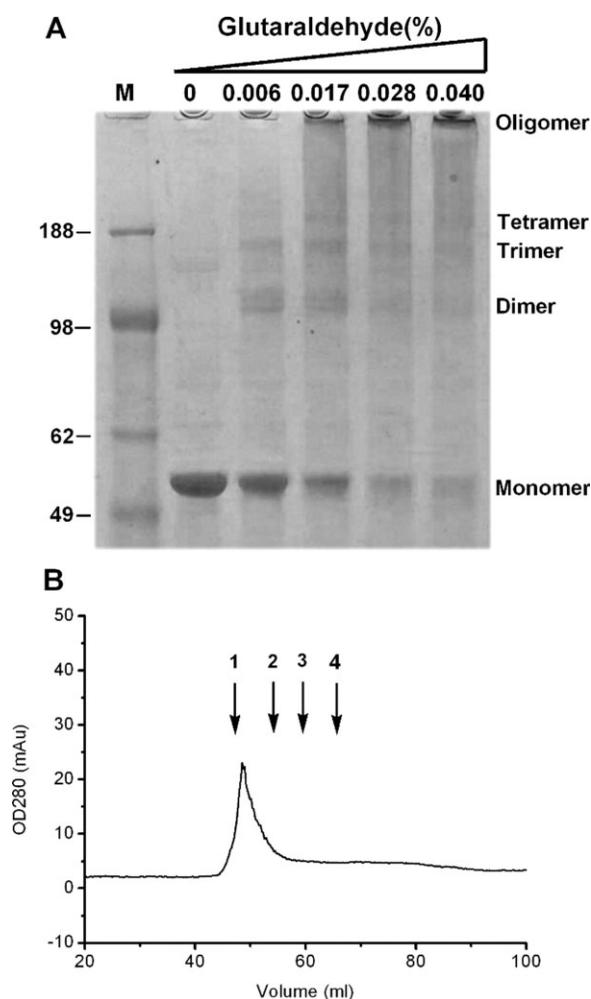
#### Oligomerization of the HCoV-OC43 nucleocapsid protein

The self-association characters of SARS-CoV and HCoV-229E N proteins has been reported previously.<sup>17,44</sup> We further characterized the oligomerization of full-length HCoV-OC43 N proteins using crosslinking assays. As shown in Figure 5(A), HCoV-OC43 N protein was able to form dimers, trimers, tetramers, and higher-order multimers as the glutaraldehyde concentrations was increased. To evaluate whether electrostatic interactions stabilize the compact conformations of higher-order multimers, the oligomeric behaviors of N protein at different ionic strengths were studied. Briefly, the purified full-length N proteins, which were in a buffer containing sodium chloride, were cross-linked by 0.04% glutaraldehyde. In addition, we analyzed the oligomerization of N protein using analytical gel filtration chromatography [Fig. 5(B)]. N protein eluted with a retention volume of  $\sim 49.5$  mL, corresponding to an apparent molecular weight of 621 kDa, based on the standard curve. This implies an oligomer containing  $\sim 12$  molecules.

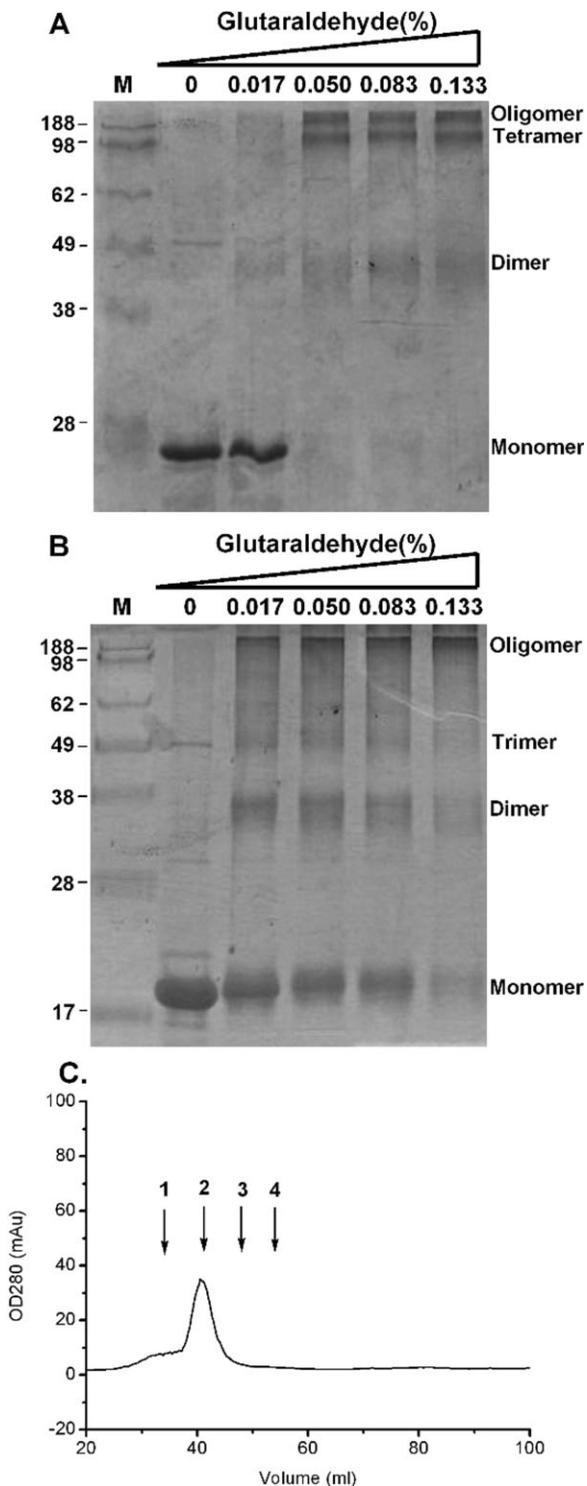
#### Identification of the oligomerization region of HCoV-OC43 nucleocapsid protein

To determine the regions that are of greatest importance for HCoV-OC43 N protein oligomerization, several regions of N protein were purified and tested using crosslinking assays and analytical gel filtration. As shown in Figure 6(A), crosslinking studies of N233-448 detected dimers, tetramers, and oligomers. A shorter form of the protein, N301-448, also formed dimers, trimers, and oligomers [Fig. 6(B)]. Gel filtration analysis showed a peak of N301-448 at high molecular weight (225 kDa), consistent with a  $\sim 12$ -mer [Fig. 6(C)]. The crosslinking assay was also conducted to investigate oligomerization of the N-terminal

(N1-173) and middle segment (N174-232) of N protein. Both truncated N proteins remained as monomers under these experimental conditions [Fig. 7], suggesting that these regions are not involved in the oligomerization. Our data showed that the 148 amino



**Figure 5.** (A) Crosslinking assay of full-length N protein in the presence of various concentrations of glutaraldehyde. The concentration of the protein was  $2 \mu M$ . (B) Gel filtration chromatography-based analysis (Superdex 300 XK16/70) of N protein dissolved in 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS. The concentration of the target protein was  $3 \mu M$ . Standard proteins of 669 kDa, 158 kDa, 67 kDa, and 43 kDa are denoted as 1, 2, 3, and 4, respectively.

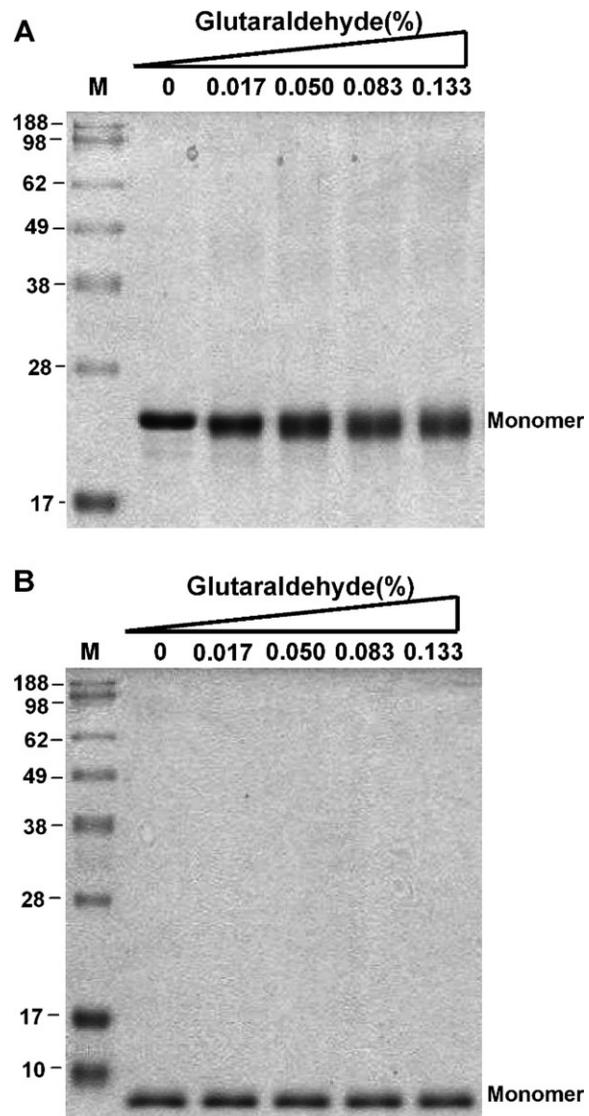


**Figure 6.** (A) Crosslinking assay of N233–448 in the presence of glutaraldehyde at various concentrations. The concentration of target protein was 2  $\mu$ M. (B) Crosslinking assay of N301–448 in the presence of the indicated concentrations of glutaraldehyde. The concentration of target protein was 3  $\mu$ M. (C) Gel filtration chromatography based analysis (Superdex 200 XK16/70) of N301–448 buffered by 50 mM Tris-HCl with 150 mM NaCl and 0.1% CHAPS at pH 7.3. The concentration of target protein was 5  $\mu$ M. The protein markers of 669 kDa, 223 kDa, 67 kDa, and 43 kDa are denoted as 1, 2, 3, and 4, respectively.

acids at the C-terminus were the oligomerization domain of HCoV-OC43 N protein.

## Discussion

We have previously analyzed the full-length SARS-CoV and HCoV-229E N proteins by CD and found them to be relatively disordered, showing nearly 50% disordered structures.<sup>17</sup> Here, we showed that the HCoV-OC43 N protein possesses a more ordered structure than other human coronavirus N proteins, including those of SARS and 229E, which feature 35%  $\alpha$ -helix content (Supporting Information Figure S2). Unlike SARS-CoV and HCoV-229E N proteins, the recombinant HCoV-OC43 N protein is highly resistant to proteolysis.<sup>17</sup> According to previous heat and urea



**Figure 7.** (A) Crosslinking assay of N1–173 in the presence glutaraldehyde at various concentrations. The concentration of target protein was 2  $\mu$ M. (B) Crosslinking assay of N174–232 in the presence of the indicated concentrations of glutaraldehyde. The concentration of target protein was 1.5  $\mu$ M.

denaturation analyses, the  $T_m$  and  $C_m$  values of SARS-CoV N protein are 38°C and 2.77 M, respectively, implying a low stability of the SARS-CoV N protein.<sup>45</sup> In this work, the  $T_m$  and  $C_m$  of the HCoV-OC43 N protein were found to be 52°C and 4 M, respectively. Our results suggest that the HCoV-OC43 N protein is more stable than the SARS-CoV N protein.<sup>45</sup>

In this study, we have shown that the recombinant HCoV-OC43 N protein expressed in *E. coli* binds to the intergenic sequence of the RNA genomic leader. This binding demonstrates that the obtained recombinant N protein was properly folded. We examined the binding of recombinant HCoV-OC43 protein for nucleic acids. Our kinetic data showed that HCoV-OC43 N protein binds preferentially to single-stranded RNA over the corresponding single-stranded DNA. The decreased affinity for DNA may be due to two factors. First, the 2'-hydroxyl group of ribose in RNA may form hydrogen bonds with residues of the N protein. Second, in DNA, the methyl group on the C5 of thymine may create a steric clash with residues within the RNA-binding groove. The Hiscox group suggested that phosphorylation of the N protein determined the recognition of virus RNA because phosphorylation may alter the conformation of N protein, and thus affect its RNA-binding sites.<sup>12,33,34</sup> In addition, Mohandas *et al.* have reported that dephosphorylation of MHV N protein by cellular phosphoprotein phosphatase may facilitate the infectious process.<sup>46</sup> However, the role of phosphorylation in the RNA-binding sites of HCoV-OC43 N protein was not investigated in our study because the N protein was expressed in *E. coli*. It would be interesting to study the effects of phosphorylation on the conformation and RNA-binding properties of HCoV-OC43 N protein in the future.

Previous studies have reported that the N-terminus of SARS-CoV N protein provides a scaffold for RNA binding.<sup>29,36,47</sup> X-ray analysis revealed that the fold of the N-terminal domain of N protein is essentially conserved across the various CoV groups. It has a U-shaped structure with the two arms rich in basic residues and the flexible loops well-ordered around the  $\beta$ -sheet core of the N-terminal domain.<sup>26,36</sup> Spencer and Hiscox found that the N-terminal region of IBV protein facilitates long-range, nonspecific interactions between N protein and viral RNA, thus leading to the formation of the ribonucleocapsid via a lure and lock mechanism.<sup>29</sup> Similar to the N-terminal domain of SARS-CoV and IBV N proteins, the N-terminal domain of HCoV-OC43 N protein contains many positively charged residues and has been shown by SPR to be responsible for RNA binding. In addition to the N-terminal domain (N1–173), two highly positively charged portions of the central linker region, N174–232 and N233–300, also show RNA-binding activity. Given their high pI values, these RNA-binding regions may rely on electrostatic interactions to interact with RNA. Compared with N1–173 and N233–300, N174–

232 has lower RNA-binding affinity and, therefore, appears to play an auxiliary role in RNA binding, despite the presence of a serine/arginine motif (SR-rich motif) that is usually involved in RNA recognition.<sup>48</sup> In a previous study, the SR-rich motif of MHV N protein was shown to be the RNA-binding region.<sup>28,42</sup> In addition, the C-terminal domain of SARS-CoV N protein has been proposed to bind nucleic acids and to participate in genomic RNA-binding.<sup>25,49</sup> However, our studies suggest that the C-terminal domain (N301–448) of the HCoV-OC43 N protein does not bind RNA.

A previous study showed that SARS-CoV N protein forms high-order oligomers through the action of its C-terminal 138 amino acids. These oligomers exist predominantly as dimers.<sup>44,50</sup> The N-terminal domain allows for the association of dimers to form higher-order oligomers. Crystal structures of the C-terminal domain of SARS-CoV and IBV N proteins show a similar general polypeptide fold, which strongly suggests that the dimerized N protein is the functional unit *in vivo* for the four groups of coronaviruses.<sup>27,35</sup> The crystal structure of the C-terminal domain shows a tightly intertwined, twofold symmetric C-terminal domain dimer with a  $\beta$ -hairpin ( $\beta_1$  and  $\beta_2$ ) from one subunit extending into the cavity of the opposite subunit; this forms an antiparallel  $\beta$ -sheet with hydrogen bonds across the dimer interface. Luo *et al.*<sup>49</sup> have reported that this dimer association does not involve electrostatic forces. These authors have further suggested that the SR-rich motif of the SARS-CoV N protein binds to the central region and thereby triggers the multimerization of dimers.<sup>51</sup> In addition, Fan *et al.* showed that the N-terminal region of the IBV N protein is involved in the oligomerization of N protein.<sup>26</sup> Here, we analyzed the self-association properties of the HCoV-OC43 N protein and found it to form oligomers. We identified the C-terminal 148 amino acids of HCoV-OC43 N protein to be the oligomerization domain. When expressed in isolation, this region of the C-terminus formed oligomers at only micromolar concentrations, and this oligomerization occurred in the absence of the N-terminal domain. The other truncated constructs of HCoV-OC43 N protein, N1–173 and N174–232, were found to remain as monomers under all of the tested experimental conditions, suggesting that these two regions are not involved in oligomerization during ribonucleocapsid formation.

In summary, we found the HCoV-OC43 N protein exhibited specific properties compared with other CoVs. For example, HCoV-OC43 N protein is more highly ordered and stabler than the SARS-CoV protein previously studied.<sup>17,45</sup> We also showed that HCoV-OC43 N protein is composed of three RNA-binding regions lying within the N-terminal region (residues 1–173) and central-linker region (residues 174–232 and 233–300) that bind to RNA in a cooperative manner. Although the C-terminal region (N301–448) lacks

RNA-binding activity, this region plays an important role in the oligomerization of the HCoV-OC43 N protein. This study may benefit the development of drugs to disrupt the binding of viral N protein with RNA and viral assembly.

### Materials and Methods

Drugs and reagents were purchased from Sigma Chemical Co. All oligoribonucleotides and oligodeoxyribonucleotides were synthesized using an automated DNA synthesizer and purified by gel electrophoresis. The biotin-linked oligomers were synthesized by incorporating the biotin synthon at the 5'-end of the oligomers; oligomers were then immobilized to the streptavidin-coated biosensor chip used for the surface plasmon resonance (SPR) experiments.

### Expression and purification of the full-length and truncated N proteins

The templates for the HCoV-OC43 N protein were provided by the Institute of Biological Chemistry, Academia Sinica (Taipei, Taiwan). To generate both full-length and truncated forms of the recombinant N proteins, the N protein gene was amplified by polymerase chain reaction (PCR) from plasmid pGENT using various primers. The PCR products were digested with *NdeI* and *XhoI*, and the DNA fragments were cloned into pET28a (Novagen) using T4 ligase (NEB). Bacteria transformed with the resultant plasmid were grown in culture. Protein expression was induced by supplementing the culture medium with IPTG to 1 mM followed by incubation at 10°C for 24 h. After harvesting the bacteria by centrifugation (3500g, 30 min, 4°C), the bacterial pellets were lysed with lysis buffer (50 mM Tris-buffered solution, pH 7.3, 150 mM NaCl, 0.1% CHAPS, and 15 mM imidazole). Soluble proteins were obtained from the supernatant following centrifugation (15,000 rpm, 30 min, 4°C) to remove the precipitate. Full-length and truncated N proteins carrying a His6-tag at their N-termini were purified using a Ni-NTA column (Novagen) with an elution gradient ranging from 15 to 300 mM imidazole. The pure fractions were collected and dialyzed against low-salt buffer. Since the N protein is positively-charged, it was further purified by SP cation exchange chromatography using a gradient from 0.05M to 1.5M NaCl in 50 mM Tris at pH 7.3, 0.1% CHAPS. The protein concentrations were determined using the Bradford method with Bio-Rad protein assay reagents.

### Circular dichroism spectroscopy

Circular dichroism (CD) spectra were obtained using a JASCO-815 CD spectropolarimeter. Temperature was controlled by circulating water at the desired temperature in the cell jacket. Each protein were dissolved in 50 mM Tris-buffered solution, pH 7.3, 150 mM NaCl, and 0.1% CHAPS. The CD spectra were collected between 250 and 190 nm with 1 nm bandwidth at

1 nm intervals. All spectra were obtained from an average of five scans. The photomultiplier absorbance did not exceed 600 V during the analysis. CD spectra were normalized by subtraction of the background scan with buffer alone. The mean residue ellipticity,  $[\Theta]$ , was calculated based on the equation,  $[\Theta] = MRW \times \Theta \lambda / 10 \times l \times c$ , where MRW is the mean residue weight,  $\Theta \lambda$  is the measured ellipticity in millidegrees at wavelength  $\lambda$ ,  $l$  is the cuvette pathlength (0.1 cm), and  $c$  is the protein concentration in g/mL. The results were analyzed using the SELCON3 program to calculate the percentage of each type of secondary structure.<sup>52</sup> In addition,  $T_m$  was determined from the polynomial fitting of the observed curve and taken as the temperature corresponding to half denaturation of the N protein. The first derivative of absorption with respect to temperature,  $dA/dT$ , of the melting curve was computer-generated and used for determining the  $T_m$ .

### Fluorescence spectroscopy

In the urea-induced unfolding experiment, N protein was added to a final concentration of 1  $\mu$ M to buffer containing various concentrations of urea, and the samples were incubated at 25°C for 30 min. The buffer consisted of 50 mM Tris, pH 7.3, and 0.1% CHAPS. Tryptophan fluorescence measurements were made using an F-4500 fluorescence spectrophotometer (Hitachi) equipped with a cuvette with a 1 cm light path. The excitation wavelength was 288 nm, and the emission data were collected between 300 and 400 nm.

### Surface plasmon resonance binding experiments

The affinity, association, and dissociation of N proteins and RNA were measured in a BIAcore 3000A SPR instrument (Pharmacia, Uppsala, Sweden) equipped with a SensorChip SA5 from Pharmacia; the apparatus measured binding by monitoring the refractive index change of the sensor chip surface. These changes, recorded in resonance units (RU), are generally assumed to be proportional to the mass of the molecules bound to the chip. The surface was first washed three times by injecting 10  $\mu$ L of 100 mM NaCl solution with 50 mM NaOH. To control the amount of RNA (or DNA) bound to the SA chip surface, the biotinylated oligomer was immobilized manually onto the surface of a streptavidin chip until a signal of 1200 RU was achieved in the first cell. The chip surface was then washed with 10  $\mu$ L of 10 mM HCl to eliminate non-specific binding. The second flow cell was unmodified and served as a control. The appropriate N proteins were dissolved in 50 mM Tris (pH 7.3) with 150 mM NaCl and 0.1% CHAPS, and passed over the chip surface for 140 s at a flow rate of 30  $\mu$ L/min to reach equilibrium. Blank buffer solution was then passed over the chip to initiate the dissociation

reaction; this step was continued for an additional 600 s to complete the reaction. After 600 s, the surface was recovered by washing with 10  $\mu$ L of 0.1% SDS for each single-stranded RNA (or DNA). Before fitting to the 1:1 Langmuir model, binding data were corrected by subtraction of the control to account for simple refractive index differences. Sensorgrams for interactions between RNA (or DNA) and protein were analyzed using BIA evaluation software (version 3). To verify if mass transport effects may arise during the interaction between N protein and immobilized RNA, binding experiments were performed at different flow rates and different surface binding capacity. It was found that the binding interaction between N protein and RNA is not a mass transport-limited process because the derived kinetic parameters (especially the association rate constant,  $k_a$ ) are independent of the flow rate and binding capacity.

### Chemical crosslinking assay

To investigate the oligomerization features of N proteins, a chemical crosslinking experiment was performed. A series of protein solutions containing N proteins was supplemented with various concentrations of glutaraldehyde, and each reaction mixture was incubated at room temperature for 5 min. The reaction was stopped by the addition of 1M Tris, pH 7.3 (0.5%, v/v, final concentration) and incubation on ice. The sample solution was then analyzed by SDS-PAGE.

### Analytical gel filtration chromatography

Gel filtration experiments were performed using fast protein liquid chromatography (Amersham Biosciences) equilibrated with buffer containing 50 mM Tris-HCl (pH 7.3), 150 mM NaCl, and 0.1% CHAPS at a flow rate of 0.5 mL/min. Blue dextran was used to determine of the void volume ( $V_0$ ). Several proteins of known molecular weight were used as standards (thyroglobulin, 669 kDa; catalase, 232 kDa; aldolase, 158 kDa; albumin, 67 kDa; ovalbumin, 43 kDa; chymotrypsinogen A, 25 kDa and ribonuclease A, 13.7 kDa, purchased from GE Healthcare), and their elution volumes ( $V_e$ ) were determined. The standard curve was plotted as the logarithm of the molecular weight against  $V_e/V_0$  for the standard proteins.

### Acknowledgment

We thank Dr. Tai-Huang Huang (IBS, Academia Sinica) for his helpful discussions and guidance. This manuscript was edited by American Journal Expert (AJE).

### References

1. St-Jean JR, Jacomy H, Desforges M, Vabret A, Freymuth F, Talbot PJ (2004) Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J Virol* 78: 8824–8834.
2. Vabret A, Mourez T, Gouarin S, Petitjean J, Freymuth F (2003) An outbreak of coronavirus OC43 respiratory infection in Normandy. *Fr Clin Infect Dis* 36:985–989.
3. El-Sahly HM, Atmar RL, Glezen WP, Greenberg SB (2000) Spectrum of clinical illness in hospitalized patients with “common cold” virus infections. *Clin Infect Dis* 31:96–100.
4. Gagneur A, Sizun J, Vallet S, Legr MC, Picard B, Talbot PJ (2002) Coronavirus-related nosocomial viral respiratory infections in a neonatal and paediatric intensive care unit: a prospective study. *J Hosp Infect* 51:59–64.
5. Wenzel RP, Hendley JO, Davies JA, Gwaltney JM, Jr (1974) Coronavirus infections in military recruits. Three-year study with coronavirus strains OC43 and 229E. *Am Rev Respir Dis* 109:621–624.
6. Arbour N, Day R, Newcombe J, Talbot PJ (2000) Neuroinvasion by human respiratory coronaviruses. *J Virol* 74: 8913–8921.
7. Leng Q, Bentwich Z (2003) A novel coronavirus and SARS. *N Engl J Med* 349:709.
8. Hogue BG, Brian DA (1986) Structural proteins of human respiratory coronavirus OC43. *Virus Res* 5: 131–144.
9. Lai MM, Cavanagh D (1997) The molecular biology of coronaviruses. *Adv Virus Res* 48:1–100.
10. Hiscox JA (2007) RNA viruses: hijacking the dynamic nucleolus. *Nat Rev Microbiol* 5:119–127.
11. You JH, Reed ML, Dove BK, Hiscox JA (2006) Three-dimensional reconstruction of the nucleolus using metaconfocal microscopy in cells expressing the coronavirus nucleoprotein. *Adv Exp Med Biol* 581:313–318.
12. Stohlman SA, Fleming JO, Patton CD, Lai MM (1983) Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein. *Virology* 130:527–532.
13. Masters PS, Parker MM, Ricard CS, Duchala C, Frana MF, Holmes KV, Sturman LS (1990) Structure and function studies of the nucleocapsid protein of mouse hepatitis virus. *Adv Exp Med Biol* 276:239–246.
14. Masters PS, Sturman LS (1990) Background paper. Functions of the coronavirus nucleocapsid protein. *Adv Exp Med Biol* 276:235–238.
15. Pyrc K, Jebbink MF, Berkhout B, Van Der Hoek L (2004) Genome structure and transcriptional regulation of human coronavirus NL63. *Virol J* 1:7.
16. Baric RS, Nelson GW, Fleming JO, Deans RJ, Keck JG, Casteel N, Stohlman SA (1988) Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *J Virol* 62:4280–4287.
17. Tang TK, Wu MP, Chen ST, Hou MH, Hong MH, Pan FM, Yu HM, Chen JH, Yao CW, Wang AH (2005) Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses. *Proteomics* 5:925–937.
18. Hsieh PK, Chang SC, Huang CC, Lee TT, Hsiao CW, Kou YH, Chen IY, Chang CK, Huang TH, Chang MF (2005) Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol* 79:13848–13855.
19. Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P (2007) Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *J Virol* 81:548–557.
20. Du L, Zhao G, Lin Y, Chan C, He Y, Jiang S, Wu C, Jin DY, Yuen KY, Zhou Y, Zheng BJ (2008) Priming with rAAV encoding RBD of SARS-CoV S protein and boosting with RBD-specific peptides for T cell epitopes elevated humoral and cellular immune responses against SARS-CoV infection. *Vaccine* 26:1644–1651.

21. Surjit M, Liu B, Chow VT, Lal SK (2006) The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. *J Biol Chem* 281:10669–10681.
22. Zuniga S, Sola I, Moreno JL, Sabella P, Plana-Duran J, Enjuanes L (2007) Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* 357:215–227.
23. Chan KH, Cheng VC, Woo PC, Lau SK, Poon LL, Guan Y, Seto WH, Yuen KY, Peiris JS (2005) Serological responses in patients with severe acute respiratory syndrome coronavirus infection and cross-reactivity with human coronaviruses 229E, OC43, and NL63. *Clin Diagn Lab Immunol* 12:1317–1321.
24. Mourez T, Vabret A, Han Y, Dina J, Legrand L, Corbet S, Freymuth F (2007) Baculovirus expression of HCoV-OC43 nucleocapsid protein and development of a Western blot assay for detection of human antibodies against HCoV-OC43. *J Virol Methods* 139:175–180.
25. Chen CY, Chang CK, Chang YW, Sue SC, Bai HI, Riag L, Hsiao CD, Huang TH (2007) Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol* 368:1075–1086.
26. Fan H, Ooi A, Tan YW, Wang S, Fang S, Liu DX, Lescar J (2005) The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure* 13:1859–1868.
27. Jayaram H, Fan H, Bowman BR, Ooi A, Jayaram J, Collisson EW, Lescar J, Prasad BV (2006) X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol* 80:6612–6620.
28. Nelson GW, Stohman SA (1993) Localization of the RNA-binding domain of mouse hepatitis virus nucleocapsid protein. *J Gen Virol* 74 (Part 9):1975–1979.
29. Spencer KA, Hiscox JA (2006) Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region. *FEBS Lett* 580:5993–5998.
30. Chang CK, Sue SC, Yu TH, Hsieh CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang WL, Lin CH, Huang TH (2006) Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci* 13:59–72.
31. Masters PS (1992) Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. *Arch Virol* 125:141–160.
32. Peng D, Koetzner CA, McMahon T, Zhu Y, Masters PS (1995) Construction of murine coronavirus mutants containing interspecies chimeric nucleocapsid proteins. *J Virol* 69:5475–5484.
33. Spencer KA, Dee M, Britton P, Hiscox JA (2008) Role of phosphorylation clusters in the biology of the coronavirus infectious bronchitis virus nucleocapsid protein. *Virology* 370:373–381.
34. Chen H, Gill A, Dove BK, Emmett SR, Kemp CF, Ritchie MA, Dee M, Hiscox JA (2005) Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance. *J Virol* 79:1164–1179.
35. Yu IM, Oldham ML, Zhang J, Chen J (2006) Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. *J Biol Chem* 281:17134–17139.
36. Saikatendu KS, Joseph JS, Subramanian V, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P (2007) Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J Virol* 81:3913–3921.
37. Takeda M, Chang CK, Ikeya T, Guntert P, Chang YH, Hsu YL, Huang TH, Kainosho M (2008) Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol* 380:608–622.
38. Kamahora T, Soe LH, Lai MM (1989) Sequence analysis of nucleocapsid gene and leader RNA of human coronavirus OC43. *Virus Res* 12:1–9.
39. Bolen DW, Santoro MM (1988) Unfolding free energy changes determined by the linear extrapolation method. II. Incorporation of delta G degrees N-U values in a thermodynamic cycle. *Biochemistry* 27:8069–8074.
40. Wang Y, Zhang X (2000) The leader RNA of coronavirus mouse hepatitis virus contains an enhancer-like element for subgenomic mRNA transcription. *J Virol* 74:10571–10580.
41. Sawicki SG, Sawicki DL, Siddell SG (2007) A contemporary view of coronavirus transcription. *J Virol* 81:20–29.
42. Nelson GW, Stohman SA, Tahara SM (2000) High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J Gen Virol* 81:181–188.
43. Welter LM, Stohman SA, Deans RJ (1990) MHV leader RNA secondary structure affects binding to the nucleocapsid protein. *Adv Exp Med Biol* 276:247–254.
44. Yu IM, Gustafson CL, Diao J, Burgner JW, Li Z, Zhang J, Chen J (2005) Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J Biol Chem* 280:23280–23286.
45. Wang Y, Wu X, Wang Y, Li B, Zhou H, Yuan G, Fu Y, Luo Y (2004) Low stability of nucleocapsid protein in SARS virus. *Biochemistry* 43:11103–11108.
46. Mohandas DV, Dales S (1991) Endosomal association of a protein phosphatase with high dephosphorylating activity against a coronavirus nucleocapsid protein. *FEBS Lett* 282:419–424.
47. Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW, Olejniczak ET (2004) Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* 43:6059–6063.
48. Lin S, Fu XD (2007) SR proteins and related factors in alternative splicing. *Adv Exp Med Biol* 623:107–122.
49. Luo H, Chen J, Chen K, Shen X, Jiang H (2006) Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* 45:11827–11835.
50. Chang CK, Sue SC, Yu TH, Hsieh CM, Tsai CK, Chiang YC, Lee SJ, Hsiao HH, Wu WJ, Chang CF, Huang TH (2005) The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett* 579:5663–5668.
51. Luo H, Ye F, Chen K, Shen X, Jiang H (2005) SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. *Biochemistry* 44:15351–15358.
52. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 209:32–44.