

# SCIENTIFIC REPORTS



OPEN

## Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers

Yingying Cong<sup>1</sup>, Franziska Kriegenburg<sup>1</sup>, Cornelis A. M. de Haan<sup>2</sup> & Fulvio Reggiori<sup>1</sup> 

Coronaviruses (CoV) are enveloped viruses and rely on their nucleocapsid N protein to incorporate the positive-stranded genomic RNA into the virions. CoV N proteins form oligomers but the mechanism and relevance underlying their multimerization remain to be fully understood. Using *in vitro* pull-down experiments and density glycerol gradients, we found that at least 3 regions distributed over its entire length mediate the self-interaction of mouse hepatitis virus (MHV) and severe acute respiratory syndrome coronavirus (SARS-CoV) N protein. The fact that these regions can bind reciprocally between themselves provides a possible molecular basis for N protein oligomerization. Interestingly, cytoplasmic N molecules of MHV-infected cells constitutively assemble into oligomers through a process that does not require binding to genomic RNA. Based on our data, we propose a model where constitutive N protein oligomerization allows the optimal loading of the genomic viral RNA into a ribonucleoprotein complex via the presentation of multiple viral RNA binding motifs.

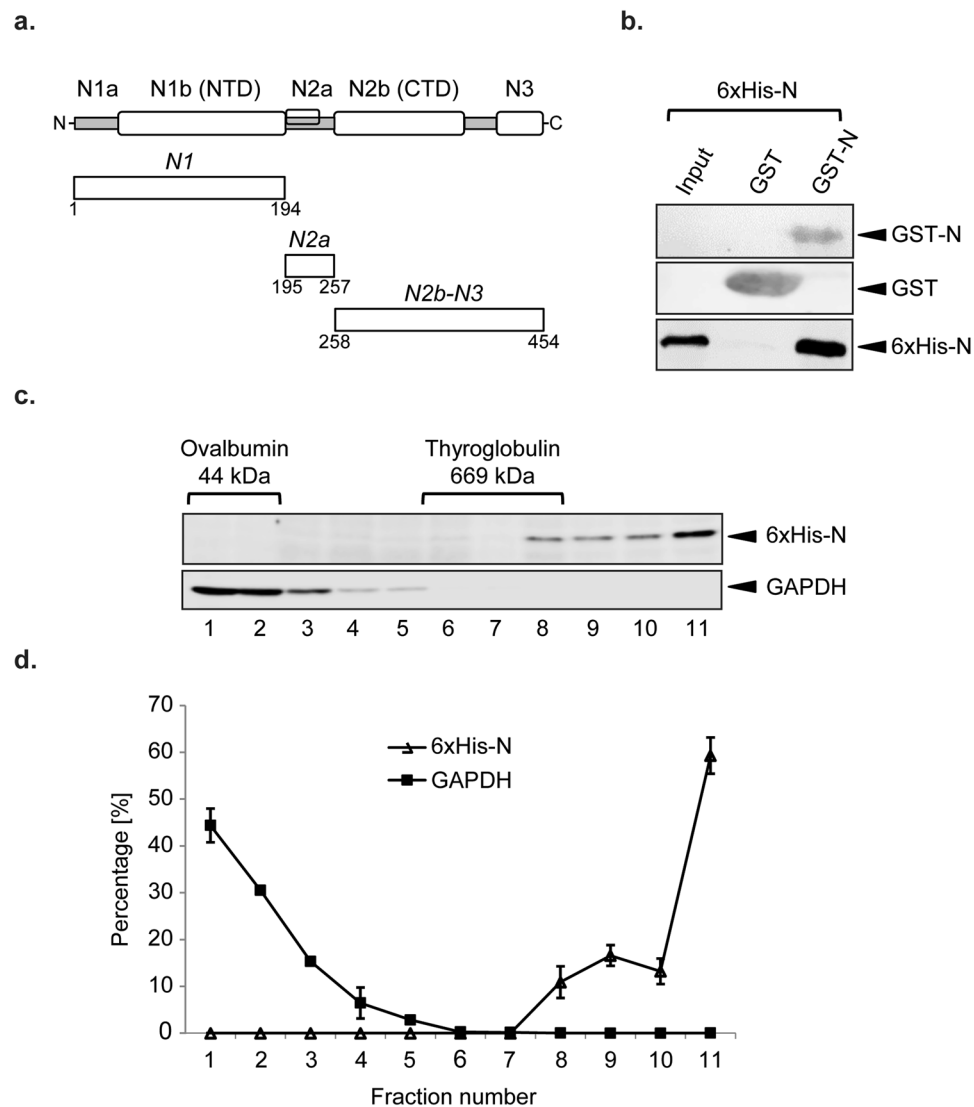
*Coronaviruses* (CoV) are enveloped positive-stranded RNA viruses and *Coronaviridae* can be subdivided into four groups based on phylogenetic clustering: *alpha*-, *beta*-, *gamma*- and *delta*-CoV<sup>1,2</sup>. Members of this virus family infect the mammalian respiratory and gastrointestinal tracts by incompletely understood mechanisms<sup>2,3</sup>. The relevance of this virus family has considerably increased due to the recent emergence of the severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), which are caused by viruses belonging to the *beta*-CoV group<sup>4,5</sup>. The Mouse Hepatitis Virus (MHV) is closely related to SARS-CoV and MERS-CoV, and considered the prototype for the investigation of the CoV life cycle<sup>6</sup>.

The proteins encoded by the CoV genomic RNA (gRNA) can be divided into two major categories. The first entails the 15 or 16 nonstructural proteins (nsp1 to nsp15/16)<sup>1,2,7</sup>, which are synthesized in the host cell and assemble into the replicase-transcriptase complexes (RTCs). RTCs are associated with and/or are embedded into double-membrane vesicles (DMVs) and convoluted membranes, which are generated during CoV infection and very likely act as replication platforms<sup>8</sup>. The second category contains the structural and accessory proteins. A minimal set of 4 structural proteins is critically required for the efficient formation of infectious virions<sup>1,2,7</sup>. Those include the envelope (E), the membrane (M), the spike (S) and the nucleocapsid (N) proteins.

The N protein is the only structural protein that associates with RTCs<sup>9–11</sup>. It binds the gRNA and it is essential for the incorporation of the virus genetic material into CoV particles<sup>12,13</sup>. Moreover, it is the major component of ribonucleoprotein complex sitting in the virion cores<sup>2,7,14</sup> and thus also plays an essential architectural role in the virus particle structural organization through a network of interactions with the gRNA, the M protein and other N molecules<sup>7,8,15</sup>. The MHV N protein has been divided into multiple domains based on genetic analyses and structural studies<sup>16–20</sup>. The two largest domains, the N-terminal domain (NTD), also N1b, and the C-terminal domain (CTD), also N2b, fold independently and have gRNA-binding properties<sup>16–18,21</sup> (Fig. 1a). These two regions are flanked by the N-terminal N1a domain, the centrally located N2a domain, and the C-terminal B spacer and N3 domain<sup>21</sup> (Fig. 1a).

It has been hypothesized that dimerization and possibly oligomerization of CoV N proteins plays an essential role in virus particle assembly<sup>21–25</sup>. The ability of SARS-CoV N protein to self-interact was first demonstrated using yeast two-hybrid and co-immunoprecipitation experiments<sup>26</sup>. The crystal structures of the N2b/CTD domain of SARS-CoV and MHV N proteins confirmed this observation, leading to the notion that N

<sup>1</sup>Department of Cell Biology, University Medical Center Groningen, University of Groningen, A. Deusinglaan 1, 9713 AV, Groningen, The Netherlands. <sup>2</sup>Virology Division, Department of Infectious Diseases & Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. Correspondence and requests for materials should be addressed to F.R. (email: [f.m.reggiori@umcg.nl](mailto:f.m.reggiori@umcg.nl))



**Figure 1.** Recombinant MHV N protein forms large oligomers. **(a)** Schematic structural organization of the MHV N protein and overview of the truncations generated in this study (modified from ref. 9). **(b)** Bacterial extract from *E. coli* expressing 6xHis-tagged MHV N protein (input) was incubated with immobilized GST or GST-N protein. Precipitated proteins were eluted in SDS-sample buffer and analyzed by western blot using an anti-His monoclonal antibody. Only part of the western blot images is shown in the figure. The GST and GST-N amount were assessed by staining the PVDF membrane with Ponceau Red. **(c)** Bacterial extract from *E. coli* expressing 6xHis-tagged MHV N protein was sedimented in a 5–20% glycerol gradient at 50,000 g for 75 min. Eleven fractions were collected and protein content analysed using antibodies against the 6xHis tag. Gradient and centrifugation conditions were assessed by sedimenting a LR7 cell extract and probing the collected fractions with a GAPDH antibody. Only part of the western blot images is shown in the figure. **(d)** Quantification of the immunoblots presented in panel (c) plus standard deviation (SD) (n = 3).

protein dimers are the basic building blocks of the ribonucleoprotein complex<sup>12, 23, 24, 27–29</sup>. Subsequently, native gel electrophoresis, size exclusion chromatography and surface plasmon resonance revealed that recombinant SARS-CoV N protein forms oligomers *in vitro*, although it appears to predominantly exist as a dimer in solution in absence of gRNA<sup>30</sup>. Size exclusion chromatography and chemical cross-linking assays were also used to unveil that the N2b/CTD domain of the SARS-CoV N protein, in particular the stretch of amino acids between positions 343–402, forms autonomously oligomers in solution<sup>31, 32</sup>. The N2b/CTD domain of the MHV N protein has also been shown to bind full length protein<sup>11</sup>. However, few other studies indicated that other parts of the N protein could also be involved in the self-interaction. In particular, the N1b/NTD and N3 domains could bind full length MHV N protein<sup>11, 19</sup> while the serine-rich (SR) region located within the N2a could be essential for SARS-CoV N protein self-interaction and oligomerization<sup>33</sup>. Nonetheless, the current most commonly accepted working model is that CoV N protein constitutively dimerizes, primarily via the N2b/CTD domain, and subsequently oligomerizes during virion assembly through a mechanism that remains unclear<sup>22, 23, 27, 29, 32, 34–36</sup>. It is also unclear whether gRNA binding influences CoV N protein oligomerization.

In this study, we confirm that recombinant MHV and SARS-CoV N protein self-interact to form large oligomers. However, we could also show that, besides N2b/CTD, several domains of the N protein are involved in this process. Analysis of different MHV and SARS-CoV N protein truncations revealed that at least three regions of these proteins cross-interact between each other in an interchangeable manner. Moreover, we show that two of these regions, i.e. N1 and N2b-N3, can oligomerize autonomously. Further, in infected cells, the MHV N protein forms oligomers already in the cytoplasm and its oligomerization does not require binding to gRNA. Altogether these findings indicate that CoV N proteins self-interact and oligomerize via discontinuous regions present in domains distributed over the entire protein to generate large supra-complexes. We hypothesize that these oligomers, which are formed constitutively, provide a larger binding surface for the gRNA, which will be thus optimally engaged at the RTCs and subsequently incorporated into forming viral particles.

## Results

**MHV N protein forms large oligomers *in vitro*.** To gain insights into the self-assembly mechanism of the N protein, we decided to use an *in vitro* approach to exclude the eventual involvement of other factors in this event. Since order-disorder and secondary structure predictions coupled with sequence alignment has highlighted that all CoV N proteins have the modular organization as MHV N protein, we used this as a model first. Hence, MHV N protein was expressed in *E. coli* as a GST fusion construct and purified using glutathione Sepharose (GSH-beads). The immobilized GST fusion protein or GST was incubated with bacterial cell extract of *E. coli* expressing recombinant 6xHis-tagged full length MHV N protein. As illustrated in Fig. 1b, recombinant 6xHis-tagged N protein specifically bound GST-N protein but not GST alone, which is in agreement with previous reports<sup>19, 23, 24, 27–29</sup> and further shows that N proteins self-interact directly.

Since it has been shown that CoV N proteins also bind non-viral RNA<sup>19, 21, 37–40</sup>, we also investigated the association between purified GST-tagged and 6xHis-tagged N proteins after RNase A treatment to exclude a role of bacterial RNA in the assayed interaction. As shown in Supplementary Fig. S1a, RNase A treatment removed large part of the RNA present in the bacterial extract and nucleic acids were not detected associated to the purified N protein even when this was not exposed to RNase A. Importantly, removal of RNA did not affect the binding between the two N protein fusions (Supplementary Fig. S1b). We concluded that the bacterial RNA does not participate in the *in vitro* interaction of MHV N protein.

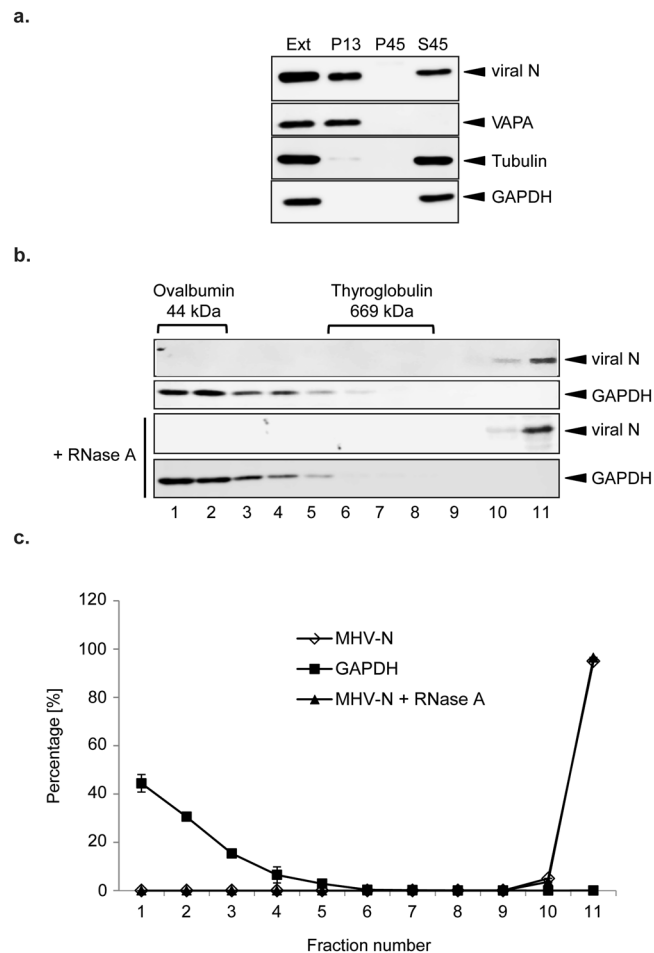
To determine whether the N-protein form homo-dimers or homo-oligomers, we sedimented recombinant 6xHis-tagged N protein on a continuous 5–20% glycerol gradient. For comparison sedimentation of the protein standards ovalbumin (44 kDa) and thyroglobulin (669 kDa) was analysed, which could be recovered in fractions 1–2 and 6–8, respectively (Fig. 1c and d). GAPDH, which is present in the cytoplasm as a monomer<sup>41</sup>, was detected in the first fractions of a sedimentation performed with a LR7 cell extract. In contrast, the 6xHis-tagged N protein was exclusively found in fractions 8–11, indicating that it forms large homo-oligomers (Fig. 1c and d).

**MHV N protein assembles in cytoplasmic oligomers over the course of an infection.** We explored whether the N protein also forms oligomers over the course of an infection. In MHV-infected cells, the N protein localizes to three distinct locations: RTCs, virions and cytoplasm<sup>2, 9, 10, 13, 42</sup>. Since the N protein is in large complexes with other viral proteins in both RTCs and virions, we decided to focus on its cytoplasmic pool and to separate it from the two other populations through differential centrifugations. We generated cell extracts (Ext) from MHV-infected LR7 cells and centrifuged them at  $15,000 \times g$  to discard the pellet P13, which includes compartments with low sedimentation rates such as whole cells or debris, nuclei and mitochondria. The resulting supernatant S13 was subsequently centrifuged at  $110,000 \times g$  to sediment organelles such as the Golgi apparatus but also virions that could have remained associated to the cell surface at the moment of the lysis<sup>43, 44</sup>. Successful fractionation of each step was followed by visualization of the membrane bound protein VAPA for P13<sup>45</sup> as well as the cytoplasmic proteins GAPDH and tubulin for S45, which also contained soluble N protein (Fig. 2a). Thus, we concluded that the cytoplasmic pool of the N protein is enriched in the S45 supernatant.

The S45 supernatant was then applied onto the same continuous 5–20% glycerol gradient employed for the analysis of the size of recombinant N protein complexes. Interestingly, cytoplasmic N protein from MHV-infected cells was exclusively detected in the last fractions of the gradient similarly to recombinant N protein, while GAPDH was found only in the low-density fractions (Fig. 2b and c). A small difference in size between the MHV N protein oligomers formed *in vitro* and *in vivo*, however, was detected. This could be due to either a slight inhibition of recombinant N protein self-interaction caused by the 6xHis tag, or a better oligomerization *in vivo* because of cellular factors such as molecular chaperones. Moreover, it cannot be excluded that there are host proteins that associate to N protein oligomers.

The S45 supernatant was also incubated with RNase A to degrade all nucleic acids (Supplementary Fig. S1c), before applying the sample onto the glycerol gradient to examine whether N protein oligomerization is influenced by binding to gRNA *in vivo*. This treatment, however, did not change the sedimentation profile of the N protein (Fig. 2b and c). Altogether these data show that the MHV N protein forms large cytoplasmic oligomers in infected cells, and that this aggregation does not depend on its binding to gRNA. Moreover, this result underlines the validity of using recombinant N protein to study its oligomerization determinants.

**Multiple domains mediate N protein oligomerization.** Previous studies have shown that the N2b/CTD domain is required for the dimerization of N proteins of different CoVs<sup>12, 23, 24, 27, 28</sup>. Our consistent finding that the N protein forms oligomers suggested that there might be several domains involved in the self-interaction. We thus generated three 6xHis-tagged truncations, i.e. N1 (which contains the NTD), N2a and N2b-N3 (which contains the CTD), which collectively cover the full length of the N protein (Fig. 1a). These constructs were expressed in *E. coli* and the resulting bacterial extracts were incubated with either immobilized GST or GST-N protein. Interestingly, all three analyzed N protein truncations specifically bound GST-N protein but not GST

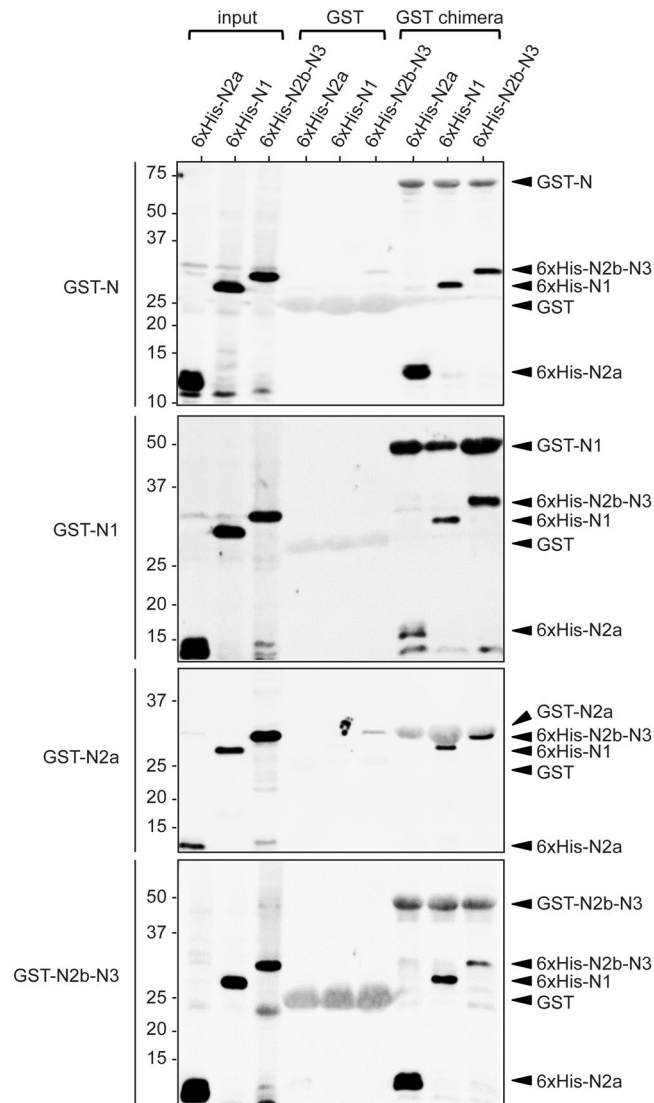


**Figure 2.** MHV N protein assembles in cytoplasmic oligomers over the course of an infection. **(a)** Cleared lysate from MHV-infected LR7 cells (Ext) was centrifuged at  $15,000 \times g$  for 10 min to obtain a pellet (P13) and a supernatant, which was subsequently centrifuged at  $110,000 \times g$  for 1 h to be separated into a pellet (P45) and a supernatant (S45). Equivalent amounts of each fraction were separated by SDS-PAGE and analysed by western blot using antibodies against MHV N protein, tubulin (cytoplasm), GAPDH (cytoplasm) and VAPA (endoplasmic reticulum). Only part of the western blot images is shown in the figure. **(b)** The S45 fraction obtained in panel (a) was mocked treated or treated with RNase A for 30 min on ice before sedimentation and analysed as in Fig. 1c. A260/280 ratios  $<0.1$  was used to assess complete hydrolysis of the nucleic acids. Only part of the western blot images is shown in the figure. **(c)** Quantification of the immunoblots presented in panel (b) plus SD ( $n = 3$ ).

alone (Fig. 3, top panel), further revealing that in addition to the reported N2b/CTD region, the N protein possesses binding domains for the self-interaction in the N1 and N2a parts as well.

Next we explored whether there was any putative redundancy in binding between the different domains of the N protein to mediate N protein oligomerization. For this, bacterial cell extracts from *E. coli* expressing 6xHis-tagged N1, N2a or N2b-N3 were incubated with immobilized GST or GST-N protein as well as the truncations GST-N1, GST-N2a and GST-N2b-N3. Interestingly, all constructs were able to specifically bind both the N1 and N2b-N3 truncations suggesting that each domain within the N protein could interact with at least two different regions from another N protein molecule (Fig. 3). Further, 6xHis-tagged N1 and 6xHis-tagged N2b-N3 displayed a binding with immobilized GST-N2a. However, there was no binding between 6xHis-tagged N2a and immobilized GST-N2a indicating that the N2a domain might participate to the oligomer formation without being one of the critical determinants. 6xHis-tagged N2b-N3 was pulled down by GST-tagged N1 protein irrespectively of RNase A treatment confirming that the studied bindings do not depend on bacterial RNA (Supplementary Fig. S1d).

To determine whether the analyzed truncations also form oligomers, we sedimented bacterial extract from *E. coli* expressing 6xHis-tagged N1, N2a or N2b-N3 truncations on a continuous 5–20% glycerol gradient. As shown in Supplementary Fig. S2a,b, the 6xHis-tagged N1 and 6xHis-tagged N2b-N3 truncations are both forming oligomers. In contrast, however the 6xHis-tagged N2a sediments at lower molecular weight fractions. The presence of the 6xHis-tagged N2a fusion protein in fraction 3–4 also suggest that the N2a fragment might be able

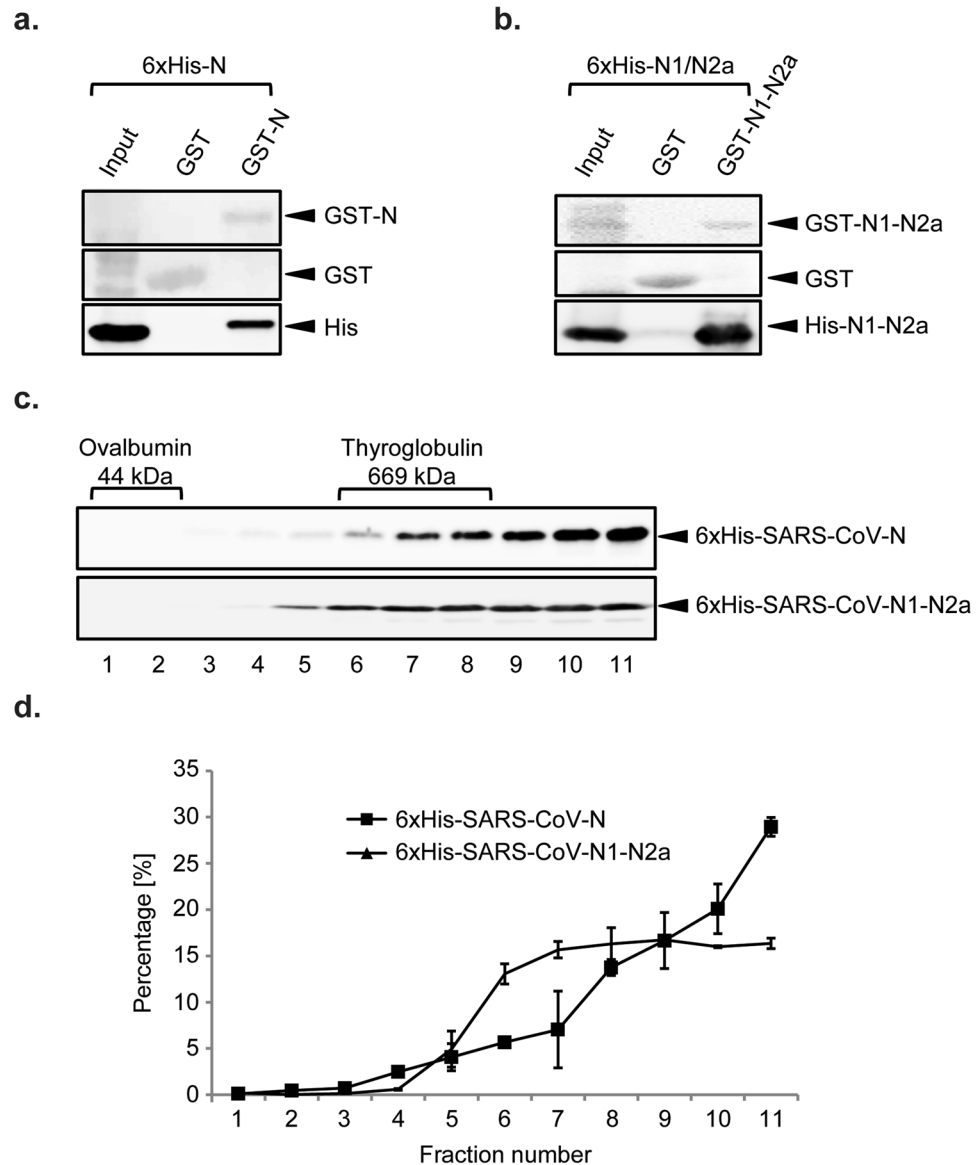


**Figure 3.** Multiple domains mediate N protein oligomerization. Bacterial extracts from *E. coli* expressing the 6xHis-tagged N1, N2a and N2b-N3 truncations were incubated with immobilized GST, GST-N protein, GST-N1, GST-N2a and GST-N2b-N3. Precipitated proteins were eluted in SDS-sample buffer and analyzed by western blot using the anti-6xHis monoclonal antibody. A260/280 ratios <0.1 indicated the absence of nucleic acids in the samples.

to multimerize through probably a weak interaction that could not be detected by pull-down experiments. From these findings we concluded that MHV N protein oligomerizes via multiple discontinuous regions.

**SARS-CoV N protein is also forming oligomers *in vitro*.** To determine whether other CoV N proteins have the same characteristic, we analyzed the SARS-CoV N protein. First, recombinant 6xHis-tagged SARS-CoV N protein was incubated with immobilized GST or GST-tagged SARS-CoV N protein. As shown in Fig. 4a, recombinant SARS-CoV N protein specifically bound to GST-SARS-CoV N protein but not GST, confirming that SARS-CoV N protein self-interacts<sup>12,23,24,27,28</sup>. Subsequently, bacterial extract from *E. coli* expressing 6xHis-tagged SARS-CoV N protein was applied onto a 5–20% glycerol gradient. The 6xHis-tagged SARS-CoV N protein was mainly detected in the late fractions of the gradient (Fig. 4c and d). These results showed that the SARS N protein, similarly to the MHV N protein, forms high molecular weight oligomers. Those of SARS-CoV N protein, however, appear to be smaller and this could be due to either the difference in size between SARS-CoV and MHV N proteins (423 amino acids versus 455) or the fact that SARS-CoV N protein forms smaller oligomers.

Previous data have shown that the N2b/CTD domain of SARS-CoV N protein domain self-interacts<sup>24,26,31,34</sup>. To explore whether the N-terminus of the SARS-CoV N protein also participates in its oligomerization, as seen for the MHV N protein, we performed pull-down experiments with recombinant 6xHis-tagged SARS-CoV-N1-N2a and immobilized GST-SARS-CoV-N1-N2a. As depicted in Fig. 4b, recombinant SARS-CoV N1-N2a protein specifically binds to GST-SARS-CoV N1-N2a protein but not GST alone, revealing that the N-terminal N1-N2a domain self-interacts. Moreover, analysis of recombinant 6xHis-tagged SARS-CoV-N1-N2a on glycerol gradients



**Figure 4.** SARS-CoV N protein is also forming oligomers. **(a)** Bacterial extract from *E. coli* expressing 6xHis-tagged SARS-CoV N protein (input) was incubated with immobilized GST or GST-SARS-CoV-N protein. Precipitated proteins were eluted in SDS-sample buffer and analyzed by western blot using an anti-His monoclonal antibody. The amounts of GST and GST-N were assessed by staining the PVDF membrane with Ponceau Red. A260/280 ratios <0.1 indicated the absence of nucleic acids in the samples. Only part of the western blot images is shown in the figure. **(b)** Bacterial extracts from *E. coli* expressing 6xHis-tagged SARS-CoV N1-N2a were processed and analyzed as in panel (a). A260/280 ratios <0.1 indicated absence of nucleic acids in the samples. Only part of the western blot images is shown in the figure. **(c)** Bacterial extracts of *E. coli* expressing 6xHis-tagged SARS N protein and N1-N2a truncation were sedimented in a 5–20% glycerol gradient at 50,000 g for 75 min. Eleven fractions were collected and protein content analyzed using antibodies against the 6xHis tag. Gradient and centrifugation conditions were assessed by sedimenting a LR7 cell extract and probing the collected fractions with a GAPDH antibody. Only part of the western blot images is shown in the figure. **(d)** Quantification of the immunoblots presented in panel (c) plus standard deviation (SD) (n = 3).

showed that this truncation forms high molecular weight oligomers (Fig. 4d and e). Similarly to MHV, however, SARS-CoV-N2a was unable to self-interact and form oligomers (Supplementary Fig. S3). Altogether our results confirm that the SARS-CoV N protein oligomerizes but the oligomerization process involves not only the CTD domain but also the N-terminal area of the N protein, which contains the NTD domain.

## Discussion

Using pull-down experiments and density glycerol gradients, our study provides additional evidence that CoV N proteins oligomerize. Several structural biology studies have indicated that SARS-CoV N protein dimerizes through the N2b/CTD domain<sup>22,24,27,32,36</sup> and it is generally accepted that the dimerization of N2b/CTD domain



serves as the basic building block for CoV ribonucleoprotein virion core formation through multimerization<sup>15, 46, 47</sup>. We and others, however, have already pointed out that the N-terminal N1b/NTD and N3 regions could also interact with full length of MHV N protein<sup>11, 19</sup>. Our data extends and completes the information acquired by these investigations as we found that at least three areas of MHV N protein, i.e. N1a-N1b, N2a and N2b-N3, mediate its self-interaction. Importantly, those domains can bind reciprocally between themselves (Fig. 3) and our findings provide a possible model for MHV N protein oligomerization, i.e. a single N molecule appears to have several binding sites that allow the association of multiple N protein units to a single oligomer. However, we cannot exclude that binding promiscuity and affinity of the various N protein domains, which we have investigated using truncated proteins, are more limited and different, respectively, in the context of the full-length protein.

Although N1a-N1b and N2b-N3 regions are able to oligomerize independently, N2a is not and therefore this domain probably only contributes to and/or reinforces MHV N protein self-interaction (Fig. 3 and Supplementary Figs S2 and S3). This is in agreement with a report indicating that the N2a domain, in particular the SR region within it, could play a role for SARS-CoV N protein self-interaction and oligomerization<sup>33</sup>. Further, the N2a domain is known to interact with the nonstructural protein nsp3 via a serine and arginine rich stretch within that domain<sup>9–11</sup>. Therefore, one can speculate that N2a is more important for the association of the N protein to the RTCs rather than providing another oligomerization platform.

Our attempts to narrow down the specific binding area within the various regions have been unsuccessful since it appears that the amino acids crucial for the N protein self-interaction are discontinuously distributed (Fig. 3). The results obtained with the N2a truncation indirectly support this notion as a short domain will possess less key binding amino acids and consequently it will interact less pronouncedly compared to longer parts such as the N1 and N2b-N3 domains. Previous efforts to localize the residues essential to mediate SARS-CoV N2b-N3 (including the CTD) multimerization identified three different regions within this domain<sup>22, 31, 34</sup>, further underscoring the notion that discontinuous binding regions, distributed over the entire N molecule, are responsible for the self-association. During the preparation of this manuscript, a cryo-EM analysis of ribonucleoprotein complexes isolated MHV virions was published and the model emerging from this work is consistent with our findings<sup>48</sup>. This study suggested that the N protein form octamers mainly via its CTD domain, which then further assembles into larger oligomeric structures that can acquire either a loose or a more compact intertwined filament shape<sup>48</sup>. In the proposed model, multiple surfaces of the N protein participate in the multimerization of the N protein octamer and this is coherent with our conclusion that several domains in the N protein mediate self-interaction.

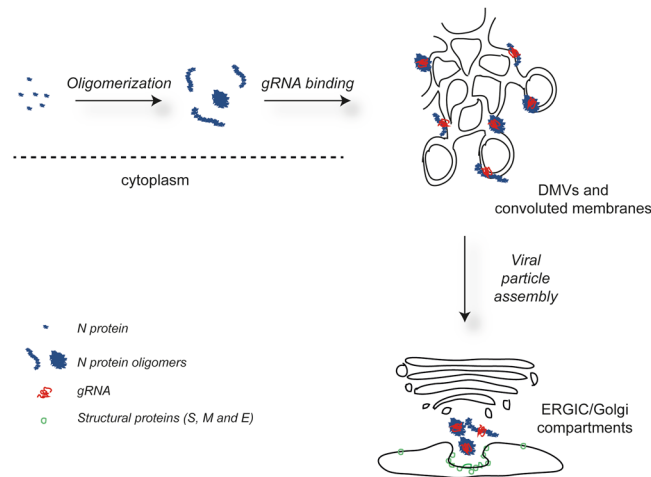
One important finding of our study is that cytoplasmic MHV N protein forms high molecular weight oligomers in infected cells. Similarly to the recombinant protein, we could not detect monomers, dimers or small multimers on glycerol gradients (Fig. 2b) suggesting that after synthesis, MHV N protein rapidly assembles into oligomers. It is easy to imagine that, based on our *in vitro* data, this is very likely also the case for the SARS-CoV N protein. Since all CoV N proteins have an identical modular organization<sup>21</sup> and have the N1b, N2a and N2b domains, oligomerization could be a characteristic that all of them possess. It has been suggested that gRNA promotes MHV N protein self-interaction because the association process was partially or largely susceptible to RNase A treatment<sup>11, 35</sup>. Another study, however, reached the opposite conclusion<sup>19</sup>, which is also supported by structural biology studies where N protein multimers have been detected and analyzed in preparations that do not contain RNA<sup>12, 22–24, 27, 29, 32, 36, 49</sup>. Our data showing that both *in vivo* and *in vitro* N protein oligomerization does not depend on its binding to gRNA are in agreement with this latter conclusion. It cannot be excluded, however, that association to gRNA could promote further N protein oligomerization. This could explain the partial discrepancy with the study showing that RNase A treatment interferes with the binding between the N protein and full-length N protein or N1b/NTD domain, but not with the N2b/CTD region<sup>11</sup>.

Which could be the relevance of constitutive N protein oligomerization? RNA chaperones are nonspecific nucleic acid binding proteins with long disordered regions that help RNA molecules to adopt its functional conformation<sup>50–52</sup>. In agreement with this notion, which has already been proposed for CoV N proteins<sup>40, 53</sup>, our hypothesis is that recruitment of already formed N protein oligomers to the RTCs at DMVs and convoluted membranes via the interaction with nsp3<sup>9, 10</sup>, allows efficient and tight loading of the exceptionally large gRNA via numerous binding sites into a ribonucleoprotein complex (Fig. 5). This RNA chaperone role of N protein oligomers would assure the efficient incorporation of the gRNA into the assembling virions, other scenarios, however, are also possible and future investigations will help to decipher the functional relevance of CoV N protein oligomerization.

## Materials and Methods

**Cell culture and virus.** LR7 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM; Cambrex Bioscience, Walkersville, MD) supplemented with 10% fetal calf serum (Bodinco Alkmaar, The Netherlands), 100 IU of penicillin/ml and 100 µg/ml of streptomycin (both from Life Technologies, Rochester, NY). Wild type MHV-A59 was propagated in LR7 cells in DMEM.

**Plasmids.** The sequences coding for either full-length MHV N protein or its truncations, i.e. N1 (amino acids 1 to 194), N2a (amino acids 195 to 257) and N2b-N3 (amino acids 258 to 454), were amplified by PCR from MHV gRNA and cloned into pET32c (EMD Millipore, Amsterdam, The Netherlands) and pGEX (GE Healthcare, Little Chalfont, United Kingdom) vectors using *Bam*HI and *Xho*I, creating the pET32c-N, pET32c-N1, pET32c-N2a, pET32c-N2b-N3, pGEX-N, pGEX-N1, pGEX-N2a and pGEX-N2b-N3 constructs. The SARS-CoV N protein coding sequence or its truncations N1-N2a (amino acids 1 to 260) and N2a (amino acids 189 to 260) were also amplified by PCR and cloned into pET32c and pGEX vectors using *Xho*I and *Not*I to create pET32c-SARS-CoV-N, pET32c-SARS-CoV-N1-N2a, pGEX-SARS-CoV-N, pGEX-SARS-CoV-N1-N2a and pGEX-SARS-CoV-N2a.



**Figure 5.** Models for the role of CoV N proteins over the course of an infection. After synthesis, CoV N proteins constitutively assemble into oligomers with loose or more compact intertwined filament shapes<sup>48</sup>, which are recruited to the RTCs localized on double-membrane vesicles (DMVs) and convoluted membranes via their interaction with nsp3. At these replication platforms, newly synthesized gRNA is engaged by N protein oligomers, which co-operate with the rest of the structural proteins to form the viral particles at the ERGIC/Golgi compartments.

**Bacterial extracts.** Transformed *Escherichia coli* BL-21 were grown in 125 ml of LB medium (0.5% yeast extract, 1% tryptone, 1% NaCl) to late exponential phase and after inducing protein expression by addition of 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside, cells were grown at 37 °C or 20 °C for 4 h or 16 h, respectively. Bacteria were harvested, resuspended in 4 ml lysis buffer (PBS, 5 mM DTT, 1 mg/ml lysozyme, 1 mM PMSE, 10% glycerol, 1% Triton X-100 and complete protease inhibitor (Roche) and lysed by two sonication rounds of 10 sec using a Branson sonicator (Danbury, Connecticut, United States). The bacterial lysates were cleared by centrifugation at  $15,000 \times g$  for 10 min at 4 °C and passed through a 0.45  $\mu$ m filter. For purification of GST fusion proteins, lysates were incubated with 125  $\mu$ l of glutathione (GSH) Sepharose (4B, GE Healthcare), which had been pre-washed in PBS. Where indicated, lysates were incubated with 40 mg/ml RNase A (Invitrogen, Carlsbad, CA) for 30 min on ice prior to addition to the GSH Sepharose. Cell extracts from bacteria expressing the 6xHis-tagged proteins were used either directly for pull-down experiments or for the purification of the fusion proteins with nickel Sepharose (6 Fast Flow, GE Healthcare) after incubation in presence or absence of 40 mg/ml RNase A for 30 min on ice. Complete hydrolysis of RNA was verified by sample analysis on an agarose gel followed by nucleic acid staining with Midori Green (GC Biotech, Netherlands) or determination of the A260/A280 ratio through measurement of A260 and A280 using a NanoDrop Spectrophotometer (Implen, Germany)<sup>54–56</sup>.

**Cell extracts.** For the preparation of cell extracts, LR7 cells grown on 10 cm dishes were mocked-treated or inoculated with MHV at a MOI of 1 and after 8 h, they were lysed by 5 min sonication in 1.2 ml of PBS buffer supplemented as described above. Supernatants were then cleared by centrifugation at  $15,000 \times g$  for 10 min at 4 °C and passed through a 0.45  $\mu$ m filter. RNase A treatments were carried out by incubating 200  $\mu$ l of cell extract with 40 mg/ml of enzyme for 30 min on ice, immediately prior to pull-downs.

**Pull-down experiments.** For the pull-down experiments, GSH-Sepharose bound GST fusion protein were incubated with 200  $\mu$ l of bacterial extract or 200  $\mu$ l of LR7 cell extracts on a rotatory wheel for 2 h at 4 °C, subsequently washed at 4 °C three times in PBS supplemented with 5 mM DTT, 10% glycerol, 1% Triton X-100 and one time in PBS buffer. Proteins bound to the Sepharose beads were eluted in 20  $\mu$ l of sample buffer by boiling and subjected to SDS-PAGE, blotted onto PVDF membranes and visualized by either membrane staining with Ponceau Red or western blot analysis using anti-6xHis antibody (HIS H8, Thermo, Waltham, MA) or anti-N protein monoclonal antibodies<sup>11</sup>. Bound primary antibodies were detected using the Alexa680-conjugated goat polyclonal anti-mouse IgG antibody (Life Technologies) and signals visualized with an Odyssey system (LI-COR, Lincoln, NE).

**Subcellular fractionation and glycerol gradient sedimentation.** Cell extracts (Ext) from MHV-infected LR7 cells were centrifuged at  $15,000 \times g$  for 10 min (4 °C) to obtain a pellet (P13) and a supernatant (S13), which was further centrifuged at  $110,000 \times g$  for 60 min (4 °C) to also get a pellet (P45) and a supernatant (S45). Proportional aliquots of Ext, P13, P45 and S45 fractions were examined by resolving them by SDS-PAGE and then by probing western blot membranes with monoclonal antibodies against MHV N protein and polyclonal antisera against GAPDH (Fitzgerald, North Acton, MA), tubulin (Sigma-Aldrich, St. Louis, MO) or VAPA (Santa Cruz, Dallas, TX).

For glycerol gradient sedimentation, 100  $\mu$ l of either the S45 fraction or bacterial extracts expressing 6xHis-tagged full-length or truncated proteins were loaded on the top of a 2,2 ml continuous 5–20% glycerol gradient in lysis buffer (w/v) prepared using the Gradient Master machine (Biocomp, New Brunswick, Canada).



After centrifugation at  $135,000 \times g$  for 75 min at 4 °C in a TLS55 rotor (Beckman Coulter, Brea, CA), 11 fractions of 200  $\mu$ l were collected from the top to the bottom of the gradient. After precipitation by addition of 20  $\mu$ l of tri-chloroacetic acid (final concentration 10%), proteins were resolved by SDS-PAGE and analyzed by western blot using antibodies against the N protein, GAPDH and the 6xHis tag. Thyroglobulin (669 kDa) and ovalbumin (44 kDa) (Bio-Rad, Berkeley, CA) were used as molecular weight protein standards to determine the gradient resolution.

## References

1. Perlman, S. & Netland, J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* **7**, 439–50 (2009).
2. Fehr, A. R. & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* **1282**, 1–23, doi:10.1007/978-1-4939-2438-7\_1 (2015).
3. Cong, Y. & Ren, X. Coronavirus entry and release in polarized epithelial cells: a review. *Rev Med Virol* **24**, 308–315, doi:10.1002/rmv.1792 (2014).
4. Shehata, M. M., Gomaa, M. R., Ali, M. A. & Kayali, G. Middle East respiratory syndrome coronavirus: a comprehensive review. *Front Med*. doi:10.1007/s11684-016-0430-6 (2016).
5. Chan, H. L., Tsui, S. K. & Sung, J. J. Coronavirus in severe acute respiratory syndrome (SARS). *Trends Mol Med* **9**, 323–325 (2003).
6. Weiss, S. R. & Leibowitz, J. L. Coronavirus pathogenesis. *Adv Virus Res* **81**, 85–164, doi:10.1016/b978-0-12-385885-6.00009-2 (2011).
7. Hagemeijer, M. C. & de Haan, C. A. Studying the dynamics of coronavirus replicative structures. *Methods Mol Biol* **1282**, 261–269, doi:10.1007/978-1-4939-2438-7\_22 (2015).
8. Knoops, K. *et al.* SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol* **6**, e226, doi:10.1371/journal.pbio.0060226 (2008).
9. Hurst, K. R., Koetzner, C. A. & Masters, P. S. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J Virol* **87**, 9159–9172, doi:10.1128/jvi.01275-13 (2013).
10. Hurst, K. R., Ye, R., Goebel, S. J., Jayaraman, P. & Masters, P. S. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *J Virol* **84**, 10276–10288, doi:10.1128/jvi.01287-10 (2010).
11. Verheije, M. H. *et al.* The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. *J Virol* **84**, 11575–11579, doi:10.1128/jvi.00569-10 (2010).
12. Ma, Y. *et al.* Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus. *Prot Cell* **1**, 688–697, doi:10.1007/s13238-010-0079-x (2010).
13. Kuo, L., Koetzner, C. A. & Masters, P. S. A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virol* **494**, 100–107, doi:10.1016/j.virol.2016.04.009 (2016).
14. Barcena, M. *et al.* Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirus. *Pr Natl Acad Sci USA* **106**, 582–587, doi:10.1073/pnas.0805270106 (2009).
15. McBride, R., van Zyl, M. & Fielding, B. C. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018, doi:10.3390/v6082991 (2014).
16. Koetzner, C. A., Parker, M. M., Ricard, C. S., Sturman, L. S. & Masters, P. S. Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. *J Virol* **66**, 1841–1848 (1992).
17. 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* **76**, 4987–4999 (2002).
18. Hurst, K. R. *et al.* A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. *J Virol* **79**, 13285–13297, doi:10.1128/jvi.79.21.13285-13297.2005 (2005).
19. Hurst, K. R., Koetzner, C. A. & Masters, P. S. Identification of *in vivo*-interacting domains of the murine coronavirus nucleocapsid protein. *J Virol* **83**, 7221–7234, doi:10.1128/jvi.00440-09 (2009).
20. Grosseohme, N. E. *et al.* Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *J Mol Biol* **394**, 544–557, doi:10.1016/j.jmb.2009.09.040 (2009).
21. Chang, C. K. *et al.* Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol* **83**, 2255–2264, doi:10.1128/jvi.02001-08 (2009).
22. Chen, C. Y. *et al.* 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, doi:10.1016/j.jmb.2007.02.069 (2007).
23. Fan, H. *et al.* The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure* **13**, 1859–1868, doi:10.1016/j.str.2005.08.021 (2005).
24. Takeda, M. *et al.* 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, doi:10.1016/j.jmb.2007.11.093 (2008).
25. Huang, Q. *et al.* Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochem* **43**, 6059–6063, doi:10.1021/bi036155b (2004).
26. Surjit, M., Liu, B., Kumar, P., Chow, V. T. & Lal, S. K. The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. *Biochem Biophys Res Comm* **317**, 1030–1036, doi:10.1016/j.bbrc.2004.03.154 (2004).
27. Yu, I. M., Oldham, M. L., Zhang, J. & Chen, J. 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, doi:10.1074/jbc.M602107200 (2006).
28. Saikatendu, K. S. *et al.* Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. *J Virol* **81**, 3913–3921, doi:10.1128/jvi.02236-06 (2007).
29. Jayaram, H. *et al.* X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol* **80**, 6612–6620, doi:10.1128/jvi.00157-06 (2006).
30. Luo, H. *et al.* *In vitro* biochemical and thermodynamic characterization of nucleocapsid protein of SARS. *Biophys Chem* **112**, 15–25, doi:10.1016/j.bpc.2004.06.008 (2004).
31. Luo, H., Chen, J., Chen, K., Shen, X. & Jiang, H. Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochem* **45**, 11827–11835, doi:10.1021/bi0609319 (2006).
32. Chang, C. K., Chen, C. M., Chiang, M. H., Hsu, Y. L. & Huang, T. H. Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. *PLoS One* **8**, e65045, doi:10.1371/journal.pone.0065045 (2013).
33. He, R. *et al.* Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Comm* **316**, 476–483, doi:10.1016/j.bbrc.2004.02.074 (2004).
34. Yu, I. M. *et al.* Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J Biol Chem* **280**, 23280–23286, doi:10.1074/jbc.M501015200 (2005).
35. Narayanan, K., Kim, K. H. & Makino, S. Characterization of N protein self-association in coronavirus ribonucleoprotein complexes. *Virus Res* **98**, 131–140 (2003).

36. Chang, C. K. *et al.* The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett* **579**, 5663–5668, doi:10.1016/j.febslet.2005.09.038 (2005).
37. Spencer, K. A. & Hiscox, J. A. Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region. *FEBS Lett* **580**, 5993–5998, doi:10.1016/j.febslet.2006.09.052 (2006).
38. Nelson, G. W., Stohlman, S. A. & Tahara, S. M. High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J Gen Virol* **81**, 181–188, doi:10.1099/0022-1317-81-1-181 (2000).
39. Kalodimos, C. G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389, doi:10.1126/science.1097064 (2004).
40. Zuniga, S. *et al.* Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* **357**, 215–227, doi:10.1016/j.virology.2006.07.046 (2007).
41. Hahn, R. G. Early detection of the TUR syndrome by marking the irrigating fluid with 1% ethanol. *Acta Anaesthesiol Scand* **33**, 146–151 (1989).
42. Kuo, L., Hurst-Hess, K. R., Koetzner, C. A. & Masters, P. S. Analyses of Coronavirus Assembly Interactions with Interspecies Membrane and Nucleocapsid Protein Chimeras. *J Virol* **90**, 4357–4368, doi:10.1128/jvi.03212-15 (2016).
43. Ujike, M., Huang, C., Shirato, K., Makino, S. & Taguchi, F. The contribution of the cytoplasmic retrieval signal of severe acute respiratory syndrome coronavirus to intracellular accumulation of S proteins and incorporation of S protein into virus-like particles. *J Gen Virol*, doi:10.1099/jgv.0.000494 (2016).
44. Klumperman, J. *et al.* Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J Virol* **68**, 6523–6534 (1994).
45. Alpy, F. *et al.* STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. *J Cell Sci* **126**, 5500–5512, doi:10.1242/jcs.139295 (2013).
46. Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D. & Huang, T. H. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral res* **103**, 39–50, doi:10.1016/j.antiviral.2013.12.009 (2014).
47. Surjit, M. & Lal, S. K. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. *Infect Genet Evol* **8**, 397–405, doi:10.1016/j.meegid.2007.07.004 (2008).
48. Gui, M. *et al.* Electron microscopy studies of the coronavirus ribonucleoprotein complex. *Prot Cell* **8**, 219–224, doi:10.1007/s13238-016-0352-8 (2017).
49. Renner, M. *et al.* Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *Elife* **5**, e12627, doi:10.7554/eLife.12627 (2016).
50. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nature reviews. Mol Cell Biol* **5**, 908–919, doi:10.1038/nrm1497 (2004).
51. Cristofari, G. & Darlix, J. L. The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Res Mol Biol* **72**, 223–268 (2002).
52. Lorsch, J. R. RNA chaperones exist and DEAD box proteins get a life. *Cell* **109**, 797–800 (2002).
53. Zuniga, S. *et al.* Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol* **84**, 2169–2175, doi:10.1128/jvi.02011-09 (2010).
54. Gross, I., Hohenberg, H. & Krausslich, H. G. *In vitro* assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur J Biochem* **249**, 592–600 (1997).
55. Hatton, T., Zhou, S. & Standing, D. N. RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. *J Virol* **66**, 5232–5241 (1992).
56. Kang, J., Lee, M. S. & Gorenstein, D. G. Application of RNase in the purification of RNA-binding proteins. *Analyt Biochem* **365**, 147–148, doi:10.1016/j.ab.2007.03.003 (2007).

## Acknowledgements

The authors thank Stuart Siddell (University of Bristol) for the anti-N protein monoclonal antibody. F.R. is supported by ALW Open Program (822.02.014), DFG-NWO cooperation (DN82-303), SNF Sinergia (CRSII3\_154421) and ZonMW VICI (016.130.606) grants. Y.C. is supported by a Chinese Scholarship Council PhD fellowship.

## Author Contributions

Y.C., F.K. and F.R. designed the study, Y.C. performed all the experiments, quantifications and prepared the manuscript with contributions from all the authors. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-06062-w

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017