Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain

Milan Surjit, a Boping Liu, b Purnima Kumar, a Vincent T.K. Chow, b and Sunil K. Lal a, *

a Virology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Road, New Delhi 110067, India
b Microbiology Department, National University of Singapore, Kent Ridge, Singapore 117597, Singapore

Received 1 March 2004

Abstract

Severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) caused a severe outbreak in several regions of the world in 2003. The virus is a novel coronavirus isolated from patients exhibiting atypical pneumonia and may have originated from wild animals such as civet cats in southern China. The genome of SARS-CoV is a positive-sense, single-stranded RNA whose sequence is distantly related to all known coronaviruses that infect humans and animals. Like other known coronaviruses, SARS-CoV is an enveloped virus containing three outer structural proteins, namely the membrane (M), envelope (E), and spike (S) proteins. The nucleocapsid (N) protein together with the viral RNA genome presumably form a helical core located within the viral envelope. The SARS-CoV nucleocapsid (N) protein is a 423 amino-acid, predicted phospho-protein of 46 kDa that shares little homology with other members of the coronavirus family. A short serine-rich stretch, and a putative bipartite nuclear localization signal are unique to it, thus suggesting its involvement in many important functions during the viral life cycle. In this report we have cloned the N gene of the SARS coronavirus, and studied its property of self-association to form dimers. We expressed the N protein as a fusion protein in the yeast two-hybrid system to demonstrate self-association and confirmed dimerization of the N protein from mammalian cell lysates by communoprecipitation. Furthermore, via deletion analysis, we have shown that the C-terminal 209 amino-acid region constitutes the interaction domain responsible for self-association of the N protein to form dimers.

Keywords: SARS coronavirus; Protein–protein interaction; Yeast two-hybrid system; Nucleocapsid protein

Severe acute respiratory syndrome (SARS) is a newly emerging infectious disease. The causative agent of SARS has been identified as a novel coronavirus, namely SARS-associated coronavirus (SARS-CoV) [1–3]. As of 30 June 2003, 8,447 probable SARS cases including 811 deaths have been reported to the World Health Organization (WHO) from 32 countries or regions worldwide.

The SARS viral genome comprises approximately 30,000 nucleotides, which are organized into approximately 13–15 open reading frames (ORFs), taking into consideration only those exceeding 50 amino acids in translational capacity [4,5]. Sequence comparison with corresponding ORFs of other known coronaviruses reveals a similar pattern of gene organization typical of coronaviruses [6].

The 1259-nucleotide N gene of the SARS virus resides at the 3’ end of the genome. This 46-kDa structural protein has been predicted to interact with the viral RNA, and both constitute the viral nucleocapsid. The N protein of coronaviruses is present in the icosahedral core as well as in the internal helical nucleocapsid [7]. Since no studies have hitherto been reported on the structure and function of the N protein of the SARS coronavirus, we PCR amplified and cloned the N protein of SARS-CoV, and studied its capability to form a dimer by self-association using the yeast two-hybrid system. Here we report that the N protein of the SARS coronavirus is capable of self-association and requires...
the C-terminal helix-rich 209 amino-acid region to facilitate this protein-protein interaction.

Materials and methods

Growth media, yeast strains, and plasmid constructs. All strains, plasmids, and plasmid constructs used in this study are described in Table 1. The full-length N gene of the SARS coronavirus (Singapore isolate) was PCR-amplified from a genomic construct of clone NC_004718, and cloned into the pCR-XL-TOPO vector (Invitrogen). The full-length N gene was excised from the pCR-XL-TOPO-N construct using the restriction enzymes EcoRI and CclI, and ligated into the pGADT7 vector to generate an N-terminal in-frame fusion with the GAL4 activation domain (AD). To clone N in fusion with the DNA-binding domain of GAL4, pGADT7-N was digested with NcoI and PsI, and the excised fragment was ligated into the pAS2 vector which was linearized with the same pair of restriction enzymes. HA tagged N was cloned into pSGI by digesting pGADT7-N with BglII and ligation into pSGI digested with the same enzyme. Mv-c-tagged N was produced by ligating the BamHI and ApaI fragment from pCR-XL-TOPO-N into pcDNA 3.1 vector digested with the same enzymes. All DNA manipulations were performed as described by Sambrook et al. [8]. All deletion constructs were generated by subcloning the full-length N gene as summarized in Table 1. All constructs were verified by restriction digestion and sequencing.

Yeast two-hybrid techniques. The GAL4-based two-hybrid system, kindly provided by Dr. Stephen Elledge [6], contain pAS2 (DNA binding domain vector) and pACT2 (activation domain vector), together with the yeast reporter strain Saccharomyces cerevisiae AH109 (Table 1) were employed. The host strain containing pAS2-SNF1 and pACT2-SNF4 was used as a positive control [9]. The AH109 host strain contains integrated copies of both HIS3 and lacZ reporter genes under the control of GAL4 binding sites. The AH109 yeast strain was transformed with the appropriate plasmids using the lithium acetate transformation method and plated on selective media. The full-length N protein can form a dimer

Table 1. The full-length N gene of the SARS coronavirus (Singapore isolate)was cloned into the yeast two-hybrid vectors (Table 1) resulting in an N-terminal in-frame fusion of the GAL4-DNA binding domain and the GAL4 activation domain to the N protein. In order to confirm the reading frame and expression of the N

Table 1. Yeast strains, plasmids, and recombinant plasmid constructs used in this study

| Strain/plasmid/construct | Genotype/description |
|--------------------------|----------------------|
| AH109                    | MATa, ura3-52 leu2-3,112, his3d200, trp1d1, ade2, LYS2::LexA op)-HIS3, ura3::LexA-op)-LacZ, LexA-MS2 coat (TRP1). |
| pGADT7 N                 | pCR-XL-TOPO-N cut with EcoRI and ClaI and cloned into pGADT7 cut with the same. |
| pAS2-1 N                 | pGADT7 N cut with NcoI and PsI and cloned into pAS2-1 cut with the same. |
| 1-310 N                  | pGADT7 N cut with SacI, klenowed and vector self-ligated. |
| 1-630 N                  | pGADT7 N cut with EcoRI and Nhel and fragment cloned into pGADT7 cut with EcoRI and Smal. |
| 1-830 N                  | pGADT7 N cut with HIII, fragment cloned into pGADT7 cut with EcoRI and Smal. |
| 310-1260 N               | pGADT7 N cut with SacI and Clal, fragment cloned into pGADT7 cut with Smal and Clal. |
| 630-1260 N               | pGADT7 N cut with Nhel and Clal, fragment cloned into pGADT7 cut with Smal and Clal. |
| pSGI HA N                | pGADT7 N cut with BglII and BamHI and cloned into pSGI cut with BglII. |
| pCDNA 3.1 myc N          | pCR-XL-TOPO-N cut with BamHI and ApaI and cloned into pcDNA3.1 myc cut with the same. |
protein, it was expressed in vitro using the rabbit reticulocyte-based, coupled transcription–translation system (Promega) and immunoprecipitated using anti-HA antibody (data not shown). *S. cerevisiae* AH109 (MATa trp1-901 his3 leu2-3, 112 ura3-52 ade2 gal4 gal80UR-A3::GAL-lacZ LYS2::GAL-HIS3) cells were transformed with single plasmids, or cotransformed with the GAL4 BD- and AD-vectors containing SARS-CoV N. The AH109 host strain containing pAS2-SNF1 and pACT2-SNF4 was used as positive controls [9]. AH109 contains integrated copies of both HIS3 and *lac*Z reporter genes under the control of GAL4-binding sites. The results of the two-hybrid assay are shown in Fig. 1. Single transformants used in this assay were AH109 yeast cells containing BD-N, and cells containing AD-N. Yeast host cells containing only BD- and only AD-vectors, were also used as negative controls (Fig. 1). All these transformants grew on the YPD non-selective media. The untransformed host cells were also plated as negative controls. Single transformants containing the BD-vector alone or as a fusion with N showed growth on the synthetic dextrose Trp⁻ plate (SDTrp⁻). Correspondingly, single transformants containing the AD-vector alone or as a fusion with N showed growth on the synthetic dextrose Leu⁻ plate (SDLeu⁻). The cotransformants were similarly plated on YPD and synthetic dextrose medium lacking Trp or Leu or both, and the resultant colonies were subsequently plated on His⁻ medium (SDTrp⁻ Leu⁻ His⁻) to test for His prototrophy. Growth of the cotransformants, containing the BD-N and AD-N constructs, in both SDTrp⁻ and SDLeu⁻ plates simply showed that both plasmids were present in the transformed cells. Growth of these clones on the SDTrp⁻ Leu⁻ His⁻ media showed that the transcription of the HIS3 gene was switched on by the reconstitution of the GAL4 transactivator due to a specific N–N interaction. Colonies were transferred onto nitrocellulose filters and a β-galactosidase filter assay was performed [12–14]. The cotransformants containing both the BD-N and AD-N constructs together with all positive and negative controls were tested by this assay. Results obtained from the β-galactosidase filter assay were in agreement with the results obtained from the SDHis⁻ growth experiments.

The liquid β-galactosidase assay was conducted and activity was determined with the substrate chlorophenol red-β-D-galactopyranoside (CPRG) as described previously [15,16]. The host strain AH109 alone, together with single transformants containing the AD-N and BD-N and cotransformants containing AD-/BD- without a fusion protein, BD-/AD- and single transformants with either BD-N or AD-N were tested. Negative con-

| Plasmid Constructs | YPD | SDLeu⁻ | SDTrp⁻ | SDTrp⁻/Leu⁻ | SDHis⁻ | β-Gal |
|--------------------|-----|--------|--------|-------------|--------|-------|
| AD-                |     |        |        |             |        |       |
| BD-                |     |        |        |             |        |       |
| AD- / BD-          |     |        |        |             |        |       |
| AD-N               |     |        |        |             |        |       |
| BD-N               |     |        |        |             |        |       |
| AD-N / BD-N        |     |        |        |             |        |       |
| Clone #1           |     |        |        |             |        |       |
| AD-N / BD-N        |     |        |        |             |        |       |
| Clone #2           |     |        |        |             |        |       |
| AD-1111 / BD-1112  |     |        |        |             |        |       |

Fig. 1. Yeast two-hybrid results showing full-length homotypic interactions of the N protein of SARS-CoV. YPD yeast peptone dextrose media (no selection), SDTrp⁻, SDLeu⁻, SDTrp⁻ Leu⁻, and SDHis⁻ are restrictive growth media lacking Trp, Leu, Trp Leu, and Trp Leu His, respectively. β-Galactosidase filter assay results are also shown.
trols (host untransformed cells) showed virtually no liquid β-galactosidase activity. BD-1111/AD-11112 was the positive control whereas the clones containing BD-N/AD-N were the test samples in this experiment (Fig. 2). Relative enzymatic activity was determined in five independent transformants from each group. Our results from this assay indicate a strong protein–protein interaction between AD-N and BD-N proteins.

To further confirm these interactions in a mammalian cell environment, we performed a coimmunoprecipitation assay using N protein fused with two different tags. COS-1 cells were transfected with amino-terminal HA-tagged N or carboxy-terminal myc-tagged N or both. Lysates were subsequently immunoprecipitated using respective antibodies and immunoblotted using HA antibody. Lane 2 in Fig. 3 shows a cotransfected sample immunoprecipitated using anti-myc antibody. Lane 1 shows the HA-tagged N-transfected sample immunoprecipitated using anti-HA antibody which served as a positive control. Lane 3 shows myc-tagged N-transfected sample immunoprecipitated with anti-HA antibody which served as a negative control. Lane 4 shows myc-tagged N immunoprecipitated using myc antibody. These results clearly demonstrate that the N protein is capable of homodimerization in a mammalian cellular environment. In order to form the capsid of SARS-CoV, the nucleocapsid protein has to oligomerize. Hence our data provide a functional verification to earlier predictions regarding the N protein of the SARS virus. Further studies on the molecular dissection of N were carried out to map the region of N responsible for this interaction.

Fig. 2. Liquid β-galactosidase assay results. Single and cotransformants were analyzed for a liquid β-galactosidase assay and were compared with others. Values are given in arbitrary units. The numbers above each bar represent the mean of five independent transformants.

Fig. 3. COS1 cells were transfected with HA-N,MYC-N or both, immunoprecipitated using respective antibodies, resolved by 10% SDS–PAGE and Western blotted using anti-HA antibody as described in Materials and methods. Lane 1 shows HA-N immunoprecipitated with the same antibody, lanes 2 and 3 show cotransfected cells immunoprecipitated using anti-myc or anti-HA antibody, respectively, lane 4 shows myc-N immunoprecipitated using anti-myc antibody.
A 210 amino-acid carboxy-terminal region is responsible for N dimerization

Various deletions were designed to characterize the homodimerization domain of the N protein. Thus, plasmids pGADT7 N 1-310, pGADT7 N 1-630, pGADT7 N 1-830, pGADT7 N 310-1260, and pGADT7 N 630-1260 were constructed as described in Table 1. We investigated the strength of all these interactions for the full-length and truncated N proteins by measuring the HIS3 reporter gene in the presence of 0, 5, 10, 20, 25, and 50 mM of 3-aminotriazole (AT). These results indicated the strength of the protein–protein interaction as a function of His prototrophy. Cells containing both fusions of full-length N proteins showed growth on SD His⁻ 50 mM AT plates, thereby confirming strong homodimerization.

Our two-hybrid experiments each consisted of co-transformation of one of the pGADT7 N deletions with the full-length fusion protein pGADT7 N. The results of these experiments are shown in Fig. 4. The AD 1-830, AD 1-630, and AD 1-310 constructs cotransformed with the full-length N construct (BD-N) lacked reporter gene activation. In contrast, the AD-N 310-1260 and AD-N 630-1260 constructs when tested with full-length BD-N, revealed reporter gene activity. The AD-N 310-1260 and AD-N 630-1260 when tested with BD-N (full-length) showed reporter gene activity. However, the AD 1-830, AD 1-630, and AD 1-310, when cotransformed with their corresponding full-length N construct (BD-N), showed no reporter gene activation. However, the strength of interaction between 630-1260 and BD-N was found to be lower than that of full-length N–N or AD N 310-1260 and full length N interaction. Hence, we mapped the interaction domain for the N protein dimerization to reside in C terminal one-third amino acids. Upon investigation of the predicted secondary structure model for N, we found this region to be helix-rich and contained the putative nuclear localization signal.

Self-association of the nucleocapsid protein has been observed in many viruses. This process leads to the formation of a viral capsid which protects the genome from extracellular agents. In this report, we have used two independent assays to establish a direct interaction between the nucleocapsid molecules. This dimerization process is likely to lead to multimerization of the N protein and to subsequent capsid assembly. While the

Fig. 4. Deletion analysis for isolation of the interaction domain of the N protein of SARS-CoV. Using the yeast two-hybrid approach, each of the shown deletions was tested along with a full-length N protein and scored for marker gene activity. Histidine prototrophy was studied at 5, 10, 20, 25, and 50 mM AT concentrations on SDHis⁻ media. Liquid β-galactosidase assay was conducted on the cotransformants and mean of five samples is displayed for each deletion of the N protein. The deletions divide the N protein into four regions A, B, C, and D. Numbers above each deletion bar depict amino-acids.
functional relevance of this interaction in an in vivo context is unknown, our report nonetheless clearly proves a direct physical interaction between the SARS N protein molecules. Moreover, the strong capability to self-associate was demonstrated by the N–N cotransformed yeast cells which exhibited the ability to grow at 50 mM concentration of 3AT in the yeast two-hybrid assay. However, the exact mechanism of interaction is not known. In a wide range of viruses disulfide bond mediated self-association of nucleocapsid has been observed [17–20], but the N protein of SARS is unable to do so since it does not have any cysteine as observed from its amino acid composition. Observations have been gathered from N proteins of other RNA virus genomes (e.g., equine arthritis virus and simian hemorrhagic fever virus). The disulfide bond in the nucleocapsid protein has been proposed to enhance the stability of virions: e.g., hepatitis B virus [18] and human papilloma virus [19]. Interestingly, N proteins of other corona viruses contain cysteine residues, e.g., MHV N has 2 cysteine residues. Although it is not known whether cysteine plays a role in self-association of nucleocapsid in other corona viruses, the complete absence of cysteine residues in the N protein of SARS implies that it is unique among other corona viruses. It is quite possible that the SARS virus may have adopted an alternative conformational strategy to circumvent the problem of missing cysteine. In an in silico model for secondary predictions, both N proteins of MHV and SARS show a hydrophilic, helix rich C-terminal. This distinct secondary structure may be responsible for tight intramolecular self-association; however, the exact mechanism and forces driving the described self-association of the nucleocapsid proteins need further investigation. Further, since the N protein of SARS has been predicted to play other regulatory functions [21] besides capsid assembly, dimerization could be an activating switch for regulation of protein function. Moreover, in MHV, the RNA binding domain has been found to reside in similar locations between amino acids 163–229 [20], which includes the C-terminal helix-rich region. We have identified the dimerization domain of the N protein of SARS to be localized to the C-terminal helix rich region thus suggesting a similar strategy by the SARS virus for nucleocapsid formation.

Acknowledgments

This work was supported jointly by internal funds from the International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, and a grant from the Department of Biotechnology to S.K.L. The visiting resource persons program sponsored by ASM-UNESCO and support from Microbiology Department, National University of Singapore are gratefully acknowledged. M.S. is a senior research fellow of the CSIR. Technical support from Yeo Wee Ming is gratefully acknowledged.

References

[1] J. Buttner, Evaluation of the diagnostic value of laboratory investigations, J. Clin. Chem. Clin Biochem. 15 (1977) 1–12.
[2] A.T. Fleischauer, CDC SARS Investigative Team, Outbreak of severe acute respiratory syndrome—worldwide, Morb. Mortal. Wkly. Rep. 52 (2003) 226–228.
[3] Cumulative number of SARS probable cases in Taiwan, SARS Online Information Centre, Center for Disease Control, Taiwan [Accessed July 2, 2003]. Available from: URL: http://www.cdc.gov.tw/sarsen.
[4] M.A. Marra, S.J. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S. Butterfield, J. Khattra, J.K. Asano, S.A. Barber, S.Y. Chan, A. Cloutier, S.M. Coughlin, D. Freeman, N. Girm, O.L. Griffith, S.R. Leach, M. Mayo, H. McDonald, S.B. Montgomery, P.K. Pandoh, A.S. Petrescu, A.G. Robertson, J.E. Schein, A. Siddiqui, D.E. Smailius, J.M. Stott, G.S. Yang, F. Pummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T.F. Booth, D. Bowness, M. Czb, M. Drehb, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabany, Y. Li, S. Normand, U. Stroher, G.A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R.C. Brunham, M. Krajden, M. Petrie, D.M. Skowronski, C. Upton, R.L. Roper, The genome sequence of the SARS-associated coronavirus, Science 300 (2003) 1399–1404.
[5] P.A. Rota, M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M.H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J.L. DeRisi, Q. Chen, D. Wang, D.D. Erdman, T.C. Peret, C. Burns, T.G. Kaizek, P.E. Rollin, A. Sanchez, S. Liflick, B. Holloway, J. Limor, K. McCausland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A.D. Osterhaus, C. Drosten, M.A. Pfallans, L.J. Anderson, W.J. Bellini, Characterization of a novel coronavirus associated with severe acute respiratory syndrome, Science 300 (2003) 1394–1399.
[6] M.M.C. Lai, D. Cavanagh, Adv. Virus Res. 48 (1997) 1–100.
[7] C. Risso, I.M. Anton, L. Enjuanes, J.L. Carrascosa, J. Virol. 70 (1996) 4773–4777.
[8] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[9] J.W. Harper, G.R. Adami, N. Wei, K. Keyomarsi, S.J. Elledge, Cell 75 (1993) 805–816.
[10] M.D. Rose, F. Winston, P. Hieter, Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1990.
[11] C. Bai, S.J. Elledge, Methods Enzymol. 273 (1996) 331–347.
[12] S. Tyagi, H. Korkaya, M. Zafrullah, S. Jameel, S.K. Lal, The phosphorylated form of ORF3 protein of hepatitis E virus interacts with its non-glycosylated form of the major capsid protein, ORF2, J. Biol. Chem. 277 (2002) 22759–22767.
[13] M.D. Baron, K. Forsell, Oligomerization of structural proteins of rubella virus, Virology 185 (1991) 811–819.
[18] K.S. Jeng, C.P. Hu, C.M. Chang, Differential formation of disulfide linkages in the core antigen of extracellular and intracellular hepatitis B virus core particles, J. Virol. 65 (1991) 3924–3927.

[19] M. Li, P.A. Beard, P.A. Estes, M.K. Lyon, R.L. Garcia, Intercapsomeric disulfide bonds in papilloma virus assembly and disassembly, J. Virol. 72 (1998) 2160–2167.

[20] R. He, A. Leeson, A. Andonov, Y. Li, N. Bastien, J. Cao, C. Osiowy, F. Dobie, T. Cutts, M. Ballantine, X. Li, Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein, Biochem. Biophys. Res. Commun. 311 (2003) 870–876.

[21] S.K. Wooton, D. Yoo, Homooligomerization of the Porcine reproductive and respiratory syndrome virus nucleocapsid protein and the role of disulfide linkages, J. Virol. 77 (2003) 4546–4557.