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.
Yeast expression and characterization of SARS-CoV N protein

Shishu Cao a, Haihe Wang a, Arthur Luhr a, Sek-Man Wong a,b,∗

a Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore
b Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore

Received 11 February 2005; received in revised form 10 June 2005; accepted 14 June 2005

Available online 18 July 2005

Abstract

The severe acute respiratory syndrome human coronavirus (SARS-CoV) nucleocapsid protein (N protein) is its most antigenic structural protein. The N protein gene has been cloned into a yeast expression vector pPIC9, transformed into Pichia pastoris strain GS115 and induced for expression by methanol. SDS-PAGE and Western blot showed that the N protein was expressed at a level of 3 mg/ml of culture medium. Characterization by mass spectrometry, circular dichroism and fluorescence luminescence assays showed that the expressed N protein displayed a β-sheet secondary structure in solution and it is stable in the pH range between 5.0 and 8.0. The P. pastoris-expressed N protein is believed to more closely resemble native SARS N protein than the bacterially expressed N protein.

Keywords: Pichia pastoris; SARS-CoV; N protein; Yeast expression

1. Introduction

The genome sequence of SARS coronavirus (SARS-CoV) is 29,727 nucleotides long, with at least 11 open reading frames (ORFs) (Marra et al., 2003). Like other coronaviruses, five major ORFs are typically arranged in the order of replicase (rep), spike (S), envelope (E), membrane (M) glycoproteins and nucleocapsid (N) protein (Rota et al., 2003).

During virion assembly, N protein binds to a specific packaging signal on the viral RNA, leading to the formation of a helical nucleocapsid. Subsequently, the helical nucleocapsid interacts with the M and E proteins to form a nucleocapsid complex, resulting in budding through the membrane (Siddell, 1995).

The N protein of SARS-CoV has a predicted molecular weight (MW) of about 46 kDa and a pI of 10.1. Among all structural proteins of SARS-CoV, the N protein has the highest hydrophilicity and pI. It is the only SARS-CoV protein which contains no cysteine residue. Thus, it contains no disulfide bonds. These features suggest that the N protein may be less stable than other structural proteins of SARS-CoV (Wang et al., 2004). Based on the available information from other coronaviruses, the N protein has been implicated in a variety of functions, including genomic RNA replication (Compton et al., 1987), subgenomic RNA transcription (Stohlman et al., 1988) and translation (Ulmanen et al., 1976).

In order to study SARS-CoV assembly in human cells, recombinant protein expressed from eukaryotic organisms would be a desired approach. The N protein plays an important role in replication and packaging of coronaviruses. The antigenicity of N protein is stronger than the spike protein and it is a preferred antigen for study (Liu et al., 2004). The yeast-expressed N protein may be more useful for in vitro study of virus assembly, virus replication, subgenomic RNA replication and translation of this human coronavirus.

2. Materials and methods

2.1. Yeast strains and plasmids

Pichia pastoris strain GS115 (his4), Escherichia coli strain TOP10F, plasmid pPIC9 was supplied from a Pichia Expression Kit (Invitrogen, Cat. No. 45-0366).
2.2. SARS-CoV N protein gene and construction of recombinant expression vector pPIC9-NP

Relevant SARS-CoV cDNA clones encompassing the N protein gene were kindly provided by the Genome Institute of Singapore (GIS). The full-length cDNA clone of the SARS-CoV N protein gene was obtained by polymerase chain reaction (PCR) with a pair of forward primers (5′-GGATTCTAGTCTGATATGGGACCCTCCATC-3′), which possessed an added EcoRI site (in bold) adjacent to the N-terminal coding sequence (in italic) and a reverse primer (5′-GATGAATATGGCCTGGCCGCTTATGCTGAGTTTGAATCAGC-3′) which possessed an added NotI site (in bold) adjacent to the C-terminal coding sequence (in italic). The PCR product of the N protein gene and the plasmid pPIC9 were digested with EcoRI and NotI before ligation according to protocol (Laemmli, 1970). Western blot was also carried out according to the standard protocol (Sambrook et al., 1989).

2.3. Transformation of P. pastoris by electroporation and expression of recombinant N protein

One microgram of SacI-linearized pPIC9-NP was added to an electroporation cuvette (BioRad) and mixed well with 80 μl of P. pastoris strain GS115 (His+ Mut+) (Invitrogen) suspension. Electroporation was carried out in accordance with the procedure recommended by Gene Pulser® Electro-protocol for P. pastoris at 1.5 kV, 25 μF and 400 Ω. GS115 transformants were screened and patched on minimal dextrose (MD) and minimal methanol (MM) grid plates. Transformants were confirmed on MD plates, while MM plates were used to screen for Mut+ clones.

A single colony of pPIC9-NP transformant was transferred to 25 ml of buffered minimal glycerol medium (BMGY, Invitrogen) in a 250 ml flask and grown overnight at 30 °C at 250 rpm to an absorbance of 2.0. The overnight culture was concentrated and resuspended in 100 ml buffered MM medium (BMMY, Invitrogen) to an absorbance of 1.0. Induction with 0.5% (v/v) methanol was performed for up to 5 days at 12 h intervals. After that, the culture supernatant was harvested and stored at −80 °C.

2.4. SDS-PAGE and Western blot

The culture supernatant was concentrated by using PEG 8000 and treated with 1 mM protease inhibitor PMSF (Sigma) and vortexed gently for 1 min, then mixed with 1 x SDS loading buffer and heated to 100 °C for 3 min. All SDS-PAGE gels were performed according to the standard protocol (Laemmli, 1970). Western blot was also carried out according to the standard protocol (Sambrook et al., 1989).

Subsequently, PVDF membrane was blocked in 5% non-fat milk in TBST overnight at 4 °C. The blocking buffer was removed and the membrane was probed with anti-bacterially expressed SARS-CoV N protein polyclonal antibody (an E. coli–expressed SARS-CoV N protein was produced and polyclonal antibody against it was made in the laboratory, unpublished data). Excess antibodies were removed with three washes of TBST. The membrane was incubated with diluted (1:10,000 in TBST) alkaline phosphatase-conjugated goat anti-rabbit IgG (Pierce). The blots were then visualized after incubating with BCIP/NBT (Promega).

2.5. Mass spectrometry (MS)

All spectra were acquired in positive-ion mode on an ABI 4700 mass spectrometer (Applied Biosystems) equipped with a 337 nm nitrogen laser. To obtain the amino acid sequence of the N protein, the relevant band was cut out from 12% SDS-PAGE gel after coomassie blue staining, and digestion with trypsin, and the sample was subjected to MALDI-TOF-TOF mass spectrometry.

2.6. Circular dichroism (CD) measurement

The expressed protein was separated by SDS-PAGE and the band corresponding to the MW of the N protein was excised and eluted by electrophoresis in a dialysis bag. The solution in the dialysis bag was concentrated using PEG 8000 and its concentration was estimated at A280nm. CD spectra were obtained on a Jasco J-810 spectropolarimeter employing a 1 mm light-path cuvette with 1 μM of purified N protein. All data were collected at 25 °C and spectra were obtained.

2.7. Fluorescence luminescence (FL) measurements

The FL measurements were carried out using a LS50B Luminescence Spectrometer (Perkin-Elmer) in 96 wells plate. The excitation wavelength was at A337 nm and emission data were collected between A380 nm and A480 nm. The slit widths for excitation and emission were 5 and 10 nm, respectively. All data were collected at 25 °C.

2.8. pH titrations

The N protein solution at pH 6.8 was aliquoted into 5 mM phosphate buffer at different pH values. The final protein amount was adjusted to 1 μM. The N protein solution was incubated at 25 °C for 30 min. The FL measurements were collected from each sample at 25 °C.

2.9. Effects of metal ions and anions on the thermostability of the N protein

Several major metal ions and anions were tested for their effects on the thermostability of the N protein. The metal ions and anions chosen for test with the N protein were found in...
recent literature. Thermal-induced unfolding of the N protein at different metal ion and anion concentrations was compared at pH 6.8 with samples without supplemented ions. Each sample contained a final amount of 1 μM of N protein. Data were collected at 25 °C.

3. Results

3.1. Amplification and cloning of the N protein gene

The recombinant pPIC9-NP plasmid was sequenced and compared with the original sequence (Ruan et al., 2003). The sequence obtained confirmed 100% identity. After transformation of *P. pastoris* with Sall-linearized pPIC9-NP constructs, over 500 His’ clones/μg DNA were generated on MD plates. The clones were screened for methanol utilization following the standard protocol as described in the *Pichia* expression kit manual (Invitrogen). All clones obtained were Mut+ (fast methanol utilizing) strain. Twenty Mut+ clones randomly selected were grown overnight in BMGY medium and subsequently induced for N protein expression in 100 ml medium containing 0.5% methanol in BMMY medium for 84 h. The expressed N protein was analyzed by SDS-PAGE. After concentrating the culture supernatant by 5-fold, 9 out of 20 clones showed a 46.8 kDa polypeptide band that was absent from the non-induced negative control. Among the nine clones, one particular clone that showed the highest protein expression was selected for further studies.

3.2. Expression of the SARS-CoV N protein in *P. pastoris*

The N protein was expressed by *P. pastoris*. Increasing levels of N protein expression were achieved after methanol induction (Fig. 1). SDS-PAGE analysis of the culture supernatant of *P. pastoris* containing pPIC9-NP, after incubation in methanol medium, revealed a protein band corresponding to 46.8 kDa (Fig. 1A). Western blot showed that the 46.8 kDa band reacted positively to the N protein polyclonal antibodies. No band was observed in the negative control plasmid or with plasmid containing the N protein gene but without methanol induction (Fig. 1B). It is confirmed that there was indeed N protein expression in the culture supernatant (Fig. 1B). To test the expression level of the N protein, time course experiments indicated that there was a direct relationship between band intensity and methanol induction time. The N protein expression level reached 3 mg/ml of culture medium. It showed a steady increase of the N protein expression over the methanol induction time (Fig. 1C).

3.3. Mass spectrometry (MS)

MALDI-TOF-TOF mass spectra analysis of the N protein obtained from the recombinant *P. pastoris* (Fig. 2) showed the highest peak corresponded to a fragment of the trypsin-digested N protein. The MS results showed that there were 10 fragments which covered 46% of the published SARS N protein amino acid sequence. The longest fragment (WYFYYLGTGPEA SLPYGANKE-GIVWVATEGALNTPK) exactly matched the N protein sequence from W109 to K144 (36 amino acids). The amino sequence of the N protein revealed potential phosphorylation sites.

3.4. Circular dichroism and FL spectroscopy of the N protein

The CD spectrum analysis (Sreerama and Woody, 2004) suggests that the N protein forms a β-sheet secondary structure (Fig. 3) in solution at pH 6.8. When the N protein was subjected to a high molarity of urea or GdmCl, or under non-optimal pH conditions, its conformation became unfolded. This result is in agreement with another report on the SARS-CoV N protein expressed using bacteria (Wang et al., 2004). Tryptophan (Trp) emission FL has been used to trace protein-unfolding process (Wang et al., 2004). Trp emission FL data were obtained for the unfolding process of the N protein at pHs 6.8 and 2, 10 M urea and 5 M GdmCl (Fig. 4). The maximum emission wavelength was at around 340 nm at pH 6.8. When the N protein was present in 10 M urea, or 5 M GdmCl, or in pH 2 solution, the FL intensity decreased significantly and the maximum emission wavelength shifted to red. The results indicated that the N protein in solution was unstable in the presence of a high concentration of denaturants or at a low pH (Fig. 4).
3.5. pH titrations of the N protein

The FL of the N protein reached a plateau in the range of pHs 5–8, which indicated that the N protein was relatively more stable under those conditions. The FL data also showed that the N protein became unfolded when the pH was lower than 5 or higher than 8. When the N protein was subjected to extreme pHs at 3 or 12, it became largely unfolded. Therefore, the N protein was determined to be stable between pHs 5 and 8 (Fig. 5).

3.6. Effects of metal ions and anions on the thermostability of the N protein

To assay for the thermostability of the N protein in the presence of different ions, $T_{m}$ (°C) measurements were obtained with different metal ions and anions. There was no significant difference in thermostability when the N protein was tested...
relative fluorescence intensity of SARS-CoV N protein under different pH conditions. Experiments were repeated three times, S.D. = 0.05.

Table 1

| Molarity | Ca²⁺ | Cu²⁺ | K⁺ | Mg²⁺ | PO₄³⁻ | Cl⁻ | SO₄²⁻ | No ions |
|----------|------|------|----|------|-------|-----|-------|---------|
| 0.1      | 105  | 102  | 101| 108  | 117   | 123 | 100   | 100     |
| 0.5      | 104  | 81   | 95 | 96   | 141   | 126 | 146   | 100     |
| 1.0      | 87   | 67   | 88 | 90   | 120   | 153 | 104   | 100     |

in the presence of different metal ions. However, the anions were found to enhance the thermostability of the N protein at different concentrations from 0.1 to 1.0 M (Table 1). Both chloride and sulphate ions could provide thermostability of the N protein at 1.0 and 0.5 M, respectively.

4. Discussion

To study the most antigenic SARS-CoV proteins, the N protein was found to be the preferred antigen. In this study, it was chosen for expression in a suitable system. The P. pastoris expression system is one of the promising methods for the production of expressed proteins, although at present limited viral proteins have been expressed using this system (Cereghino and Cregg, 2000). We have cloned the SARS-CoV N gene into the P. pastoris expression vector pPIC9 and achieved efficient expression of the N protein to a level of 3000 mg/l, which is higher than the formerly reported eukaryotic-expressed N protein (526 mg/l; Liu et al., 2004). In addition, different methods including MS, CD, FL were used to characterize this eukaryotic-expressed N protein. The E. coli-expressed N protein was obtained previously in our laboratory and a polyclonal antibody against this N protein was obtained from rabbit (data not shown). Using the polyclonal antibody raised against bacterially expressed N protein, the yeast-expressed N protein was detected (Fig. 1B). The N protein expression level reached its peak at 84 h post-methanol induction (Fig. 1A). This is consistent with the report by Liu et al. (2004). The MW of the expressed N protein was estimated to be 46.8 kDa. The slight difference in the MW could be due to post-translational modifications or computational errors. The MS results showed complete matching of amino acid sequence from 46% of the trypsin-digested peptides. These digested peptides, ranged in size from 0.8 to 2.4 kDa, were sequenced (Fig. 2B) and the largest peptide sequenced has 36 amino acids (Fig. 2A). This is in accordance with the published amino acid sequence of the N protein (Ruan et al., 2003).

The CD spectroscopy can determine protein secondary structure in the “far-UV” spectral region (190–250 nm). At these wavelengths, the chromophore is the peptide bond and the signal arises when it is located in a regular, folded environment. The CD results suggest the P. pastoris-expressed N protein forms a β-sheet structure and is folded under optimum pH in solution. The FL data showed that the N protein was denatured in the presence of 5 M GdmCl or 10 M urea, or at pH 2.0. These results coincide with an independent report on the characterization of the SARS-CoV N protein expressed from bacteria (Wang et al., 2004). Through the use of structural biology platforms, such as NMR or X-ray crystallography, structural differences of the N protein expressed from yeast and bacteria can be compared that can, in turn, provide insights on the conformation of the N protein.

As compared to monoclonal antibodies generated against bacterially expressed SARS-CoV N protein (He et al., 2005), the potential pool of monoclonal antibodies generated from P. pastoris-expressed SARS-CoV N protein may identify additional recognition sites that can bind to epitopes that are post-translationally modified. In conclusion, the SARS-CoV N protein can be expressed successfully for further studies using P. pastoris. In addition, the P. pastoris-expressed N protein is more likely to resemble the native SARS N protein, than the N-protein that is bacterially expressed.

Acknowledgements

We thank Professor Robert McGovern of the University of Florida at Gainesville for critical review of this manuscript and the Genome Institute of Singapore for providing the SARS-CoV cDNA clones to amplify the N protein gene. This work was supported by research Grant 154-000-212-112 from National University of Singapore (NUS) and Amersham Biosciences, Singapore. C.S.S. is a NUS research scholar.

References

Cereghino, J.L., Cregg, J.M., 2000. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. FEMS Microbiol. Rev. 24, 45–46.

Compton, S.R., Rogers, D.B., Holmes, K.V., Fertsch, D., Remenick, J., McGowan, J.J., 1987. In vitro replication of mouse hepatitis virus strain A59. J. Virol. 61, 1814–1820.

He, Q.G., Du, Q.Y., Lau, S., Manopo, I., Lu, L.Q., Chan, S.W., Fenner, B.J., Kwang, J., 2005. Characterization of monoclonal antibody
against SARS coronavirus nucleocapsid antigen and development of an antigen capture ELISA. J Virol Meth. 127, 46–53.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Liu, R.S., Yang, K.Y., Lin, J., Lin, Y.W., Zhang, Z.H., Zhang, J., Xia, N.S., 2004. High-yield expression of recombinant SARS coronavirus nucleocapsid protein in methylotrophic yeast Pichia pastoris. World J Gastroenterol. 10, 3602–3607.
Marra, M.A., Jones, S.J., Astell, CR., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khatri, J., Asano, J.K., Barber, S.A., Chan, S.Y., Ciosstian, A., Coughlin, S.M., Freeman, D., Ginn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pabolish, P.K., Petes, A.S., Robertson, A.G., Schirin, J.E., Siddiqua, A., Smalv, D.E., Stott, J.M., Yang, G.S., Plummer, F., Anderson, A., Ansell, H., Stanley, N., Bernhard, K., Boutil, T.F., Brownsea, D., Crab, M., Derbol, M., Fernando, L., Flick, K., Gaberti, M., Grey, M., Gradia, A., Jones, S., Feldman, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipp,S, G., Tyler, S., Vogt, R., Ward, D., Watson, B., Brunnham, R.C., Krakian, M., Prate, M., Skowronski, D.R., Upton, C., Roper, R.L., 2003. The genome sequence of the SARS-associated coronavirus. Science 300, 1399–1404.
Rota, P.A., Olbricht, M.S., Monroe, S.S., Nix, WA., Campagnoli, R., Icomong, J.P., Pennato, A., Bankamp, B., Mullen, K., Chen, M.H., Tong, T., Tamin, A., Luse, L., Frace, M., Dierie, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, TC., Burns, C., Kuzak, T.G., Rollin, PE., Sanchez, A., Liflick, S., Holloway, B., Limor, J., McCuslicland, K., Olsen-Rasmussen, M., Fouchier, R., Gubler, S., Osterhaus, A.D., Droste, C., Pullenslucht, M.A., Anderson, L.J., Bellin, W.J., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300, 1394–1399.
Ruan, Y.J., Wei, C.L., Ee, A.L., Vega, V.B., Thoreau, H., Su, S.T., Cha, J.M., Ng, P., Chiu, K.P., Lim, L., Zhang, T., Peng, C.K., Lin, E.O., Loo, N.M., Yee, S.L., Ng, J.F., Chen, R.E., Stanton, L.W., Long, P.M., Liu, ET, 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet 361, 1779–1785.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.
Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbour Laboratory Press, USA.