Modification of SARS-CoV S1 gene render expression in *Pichia pastoris*

Haisong Lu · Guoliang Yang · Xiaozhan Fei · Hongyan Guo · Yadi Tan · Huanchun Chen · Aizhen Guo

Received: 14 November 2005 / Accepted: 12 February 2006 © Springer Science+Business Media, LLC 2006

**Abstract**  
S1 gene fragment containing receptor-binding region was amplified by several sets of primers using Over-Lap PCR. The native S1 gene was modified at A + T abundant regions; n.t.777–1683, n.t.1041–1050, n.t.1236–1248, n.t.1317–1335, n.t.1590–1605; based on the same amino acid sequences. The modified gene was cloned into a yeast expression vector pPIC9K. The resultant plasmid pPIC9K- S1 was transformed into *Pichia pastoris* GS 115 and the protein expression was induced with methanol. SDS-PAGE confirmed that the recombinant SI was secreted in the supernatant of induced GS 115. The protein yield reached 69 mg/l. ELISA and Western blot demonstrated that the S1 could react with the convalescent sera of people infected by SARS-CoV. Furthermore, ligand blot assay showed that the recombinant S1 could react with ACE2, the receptor of SARS-CoV. The molecular mass of expressed S1 was about 70 kDa, which was higher than that of the 30 kDa expected. PNGase F deglycosylation resulted in a protein band of 30 kDa. In conclusion, the S1 gene modification rendered the high-level expression of S1 in *P. pastoris* GS 115 and the protein was secreted as a biologically active form which was hyperglycosylated.

**Keywords**  
SARS-CoV · S1 gene · *P. pastoris* · Expression · Deglycosylation · Hyperglycosylation

**Introduction**  
Severe acute respiratory syndrome (SARS), caused by a novel coronavirus SARS-CoV, is a highly contagious respiratory disease [1, 2]. The first ever SARS in the world broke out at the end of 2002 in Southern China and spread rapidly to many other countries. More than 8400 SARS cases were reported and over 900 of them died by November 27, 2003, according to WHO database [3].

SARS-CoV is tentatively classified into group IV coronavirus because of its low homology of the genomic sequence with those of other coronaviruses. However, like other enveloped viruses, SARS-CoV enters host cells by binding the Spike (S) protein to its cellular receptors [4–6]. S protein belongs to class 1 transmembrane glycoprotein and can be divided into S1 and S2 fragments by its structure although it does not experience cleavage. $S_1$ is situated in the N terminal and responsible for the binding to cellular receptor, while $S_2$ is situated in the C terminal and responsible for the fusion with cellular membrane [7]. The angiotensin-converting enzyme 2 (ACE2) was identified as the major cellular receptor of SARS-CoV, [4, 5, 8]. The receptor-binding domain of $S_1$ was determined to be within the region of 318–510 amino acids (AA) [9]. Therefore, $S_1$ plays an important role in SARS-CoV-mediated infection and transmission. Additionally, $S_1$ is a key vaccine antigen for the
control of SARS-CoV-mediated infection. However, SARS-CoV is one of the human viruses which are highly restricted for the purpose of research because of the biosafety issue. The cloning and in vitro expression of S1 may provide an alternative approach to study the virus pathogenesis and infection control.

*Pichia pastoris* is a kind of eukaryotic unicell microorganism, which grows rapidly and is easily manipulated. Like higher eukaryotic expression systems, *P. pastoris* expression system undergoes post-translational modifications such as disulfide bond formation, proteolytic processing, and glycosylation [12]. We ever cloned S1 gene fragment (n.t.777–1683) including receptor binding domain into a *P. pastoris* expression vector pPIC9K but failed to express it. More recently, S1 was successfully expressed in *P. pastoris* expression system with much lower efficiency compared to the *E. coli* system [13]. Sequence analysis of the native S1 gene excluded the code bias problem in yeast. However, it was found that many A + T repeats exist at regions of n.t.1041–1050, n.t.1236–1248, n.t.1317–1335, and n.t. 1590–1605, So-crer et al. demonstrated that the A + T abundant regions may act as a polyadenylation loop in *P. pastoris* and result in improper termination of transcription of HIV envelop glycoprotein gp120 [14]. To test the hypothesis that the A + T abundant regions may impede the expression of S1 in *P. pastoris* system, S1 gene was modified to remove the A + T repeats but keep the correct codons for translation. As a result, this modification enabled the high expression of S1 in *P. pastoris*.

**Materials and methods**

Plasmids and strains

SARS S1 gene (n.t.777–1683) (Accession AY274119) was offered by Dr. Wang Hanzhong, Wuhan Institute of Virology, Chinese Academy of Sciences. *P. Pastoris* expression vector pPIC9K and the auxotrophic strain GS115 were purchased from Invitrogen (Carlsbad, CA, USA).

| Position in S1 gene (n.t.) | Native sequence | Modified sequence |
|---------------------------|-----------------|-------------------|
| 1041–1050                 | aaA aaA atT     | aaG aaG atC       |
| 1236–1248                 | taT taT aaA     | taC taC aaG       |
| 1317–1335                 | aaT taT aaT taT| aaC taC aaC       |
| 1590–1605                 | aaT ttT aaT taT| aaC ttC aaC       |

*Note:* The bold capitals in sequence columns represent the differentiated nucleotides between the native and modified sequences where A or T were changed to G or C

S1 gene manipulation

Ten primers were designed to change the third base A or T into G or C in the trinucleotides by Over-Lap PCR based on the same meaning of the native codons in SARS-CoV S1 gene. Four regions were modified and shown in Table 1. The primers used in Over-Lap PCR were listed in Table 2.

Five fragments were firstly amplified with pairs of the primers including P1/P2, P3/P4, P5/P6, P7/P8, and P9/P10. The reaction condition included 95°C 5 min; 35 cycles for 94°C 1 min, 52°C 1 min, 72°C 50 s, and 72°C 10 min for extension. The full-length of the modified S1 gene fragment was constructed with PCR by using the previous five PCR products as the templates and the primer pair of P1/P10. The overlapped extension was programmed as 94°C 1 min, 20 cycles for 50°C 1 min, 72°C 50 s, and 72°C 10 min. Pyrobest Taq and other reagents were purchased from Takara (Dalian, China). The PCR strategy was illustrated in Fig. 1.

Construction of the expression vector

The restriction enzymes *SnaBI* and *EcoRI* designed in *P1* and *P10* were used to clone the modified S1 gene fragment into the multiple cloning site in a secretion expression vector pPIC9K. The resultant vector was designated as pPIC9K-S1. The insertion was confirmed by *SalI* and *EcoRI* digestion and sequencing.

Transformation of *P. pastoris* GS115 and identification of the recombinant

The recombinant plasmid pPIC9K-S1 was completely digested with *SalI*, transformed into competent GS115 cells previously prepared with lithium chloride (Sangon, Shanghai, China) by electroporation. 200 µl of the transformant was spread on Minimal Dextrose Medium (MD) agar plates and incubated at 30°C for 2–4 days. The single colonies were picked and transferred individually into 5 ml Yeast Extract Peptone Dextrose (YPD) broth for incubation 36 h at 30°C. The yeast cells, harvested by centrifugation at 12,000g...
The batch induction of expression was conducted with 50 ml BMGY culture of the high-expression colony when the OD$_{600}$ reached 10 as previously described. The expressed S1 protein was further concentrated by 32% saturated ammonia sulfate [16]. The S1 crude protein was finally dissolved in 1 ml TE buffer for further function analysis. The protein concentration was estimated using the formula: Protein (mg/ml) = 1.75 × A280/0.74 × A260.

Function identification of S1 protein

ELISA and Western Blot were performed to analyze the antigenic activity of S1 protein.

S$_{1}$ protein was isolated by SDS-PAGE transferred onto the nitrocellulose (NC) membrane, probed with the convalescent sera of SARS patients (provided by Dr. Huo Xixiang, Hubei province Center of Disease Control, China) and goat IgG to human immunoglobulin coupled with HRP. The NC membrane was developed with DAB [15].

Conventional ELISA was performed. S1 protein was coated on ELISA plates, overlapped with the convalescent sera of SARS patients and rabbit IgG to human immunoglobulins coupled with HRP. TMB was used to develop HRP reaction, and OD$_{630}$ was read [16].

Ligand blot assay was performed to determine the receptor binding activity of S1. The procedure was similar to Western blot, but S1 on NC membrane was sequentially probed by the SARS-CoV receptor ACE2 (expressed and purified by this lab), rabbit polyclonal antibody to human ACE2 (R & D Systems, MN, USA), and the goat IgG to rabbit immunoglobulins coupled with HRP [17].

Deglycosylation analysis of S1 protein

Deglycosylation was performed according to manufacturer’s instruction of PNGase F (New England Biolabs, MA, USA). Twenty (20) μg of S$_{1}$ protein was denatured in 1 × Denaturing Buffer (0.5% SDS, 1% β-mercaptoethanol) by boiling 10 min at 100°C and added 1/10 (v/v) of 10 × G7 Reaction Buffer (50 mM Sodium Phosphate, pH7.5) and 1/10 (v/v) 10% NP-40. PNGase F 3 μl (1–2 U/μl) was added and incubated at 37°C for 1 h. The deglycosylation was determined by

![Fig. 1 Ovelap-PCR sketch map. The red regions represent four A + T abundant regions](Image)
SDS-PAGE and protein stain with Coomassie Brilliant Blue.

Results

S1 gene modification and construction of the expression vector

We used Over-Lap PCR to modify the S1 gene fragment (n.t. 777–1683) of SARS-CoV by changing the A or T in the native sequence to G or C in four regions (n.t.1041–1050, n.t.1236–1248, n.t.1317–1335, and n.t.1590–1605). The PCR product was cloned into the expression yeast vector pPIC9K and the resultant vector was designated as pPIC9K-S1. The cloned S1 gene was sequenced and all the modifications were confirmed to be correct as designed in Table 1. No unexpected mutation was observed. Double digestion of the recombinant plasmid pPIC9K-S1 with SalI and EcoRI demonstrated the right orientation of S1 gene fragment. The integration of expression vector pPIC9 k-S1 into P. pastoris GS115 genome was confirmed by PCR using primers specific to α factor and 3′AOX1 flanked by S1 gene insert (data not shown).

Protein expression of the recombinant yeast

Methanol successfully induced the S1 protein expression in this study. The pilot induction was used to screen the positive colonies for S1 expression. Then the batch induction was performed with the high-level expression colony. The protein yield reached 69 mg/l in the supernatant of yeast culture. SDS-PAGE showed that a 70 kDa protein was expressed and secreted, whereas the S1 protein was expected to be 30 kDa. To test whether the higher molecular mass resulted from the hyperglycosylation of S1 in this yeast expression system, we deglycosylated the secreted product with PNGase F which hydrolyzes nearly all types of N-glycan chains from glycopeptides/proteins. The high molecular mass of S1 was degraded to the expected size of 30 kDa after deglycosylation in the Coomassie Brilliant Blue stained SDS-PAGE gel (Fig. 2).

The reaction of SI with antisera to SARS-CoV

Western blot demonstrated that the expressed S1 reacted with the convalescent sera of SARS-CoV patients specifically, while the yeast cells transformed with the blank expression vector could not develop any reactions. Although there was only one 30 kDa band on Coomassie Brilliant Blue-stained SDS-PAGE gel after deglycosylation of the S1 protein, Western blot analysis detected two bands, one 30 kDa band as expected and another band of about 40 kDa. The later 40 kDa band was probably the oligosaccharide cleaved by PNGase F which could be recognized by antisera to SARS-CoV but could not be stained by Coomassie Brilliant Blue (Fig. 3).

ELISA detected reaction of S1 protein with the positive sera from 2 SARS-CoV convalescent patients but there was no reaction between S1 and the negative
sera from six healthy people. The OD values of the positive sera are over three times higher than those of the negative sera, suggesting the expressed S1 could specifically react with the antisera to SARS-CoV (Fig. 4).

The reaction of S1 with the ACE2

Ligand blot assay was performed to analyze the specific interaction between S1 and its receptor ACE2. S1 protein was blotted on NC membrane and reacted sequentially with ACE2, polyclonal antibody to ACE2, and the second antibody. The ACE2 fragment (1–365AA) was expressed by this lab with the gene cloned from Vero E6 cells. The recombinant S1 reacted with ACE2 a producing a clear band on blotted NC membrane, while the mocked-recombinant yeast did not give any signals (Fig. 5).

Discussion

The methylotrophic yeast P. pastoris is an excellent protein expression system because of the simplicity in molecular genetic manipulation. P. pastoris produces foreign proteins with high efficiency either intracellularly or extracellularly. Furthermore, this yeast expression system performs many eukaryotic post-translational modifications such as processing of signal sequences, folding, glycosylation, disulfide bond formation, and proteolytic processing [12]. Expression vectors based on the methanol-inducible AOX1 promoter are integrated into the host chromosome. However, there are some disadvantages which impede the heterologous protein expression such as the codon bias, A + T composition of cDNA, glycosylation patterns, etc. High-level expression of genes in P. pastoris prefers a lower ratio of A + T and the third nucleotide of C for a trinucleotide codon. A + T rich sequence lowered transcriptional efficiency because of the transcriptional blocks [18]. For instance, AT-rich regions such as 5’-ATTATTTATAAT-3’ in the gene of HIV-1 gp120 resulted in premature transcriptional termination and the production of truncated mRNA. When G + C content was increased by synthesizing the gene, this transcriptional problem was overcome giving rise to full-length mRNA [14].

In our previous study, we failed to detect expression of SARS-COV S1 gene fragment (n.t 777–1683) which includes the receptor-binding region (n.t. 954–1530) [9]. Later we realized that this fragment had abundant AT repeats and some of AT-rich regions even span more than 7 nucleotides. Therefore it was assumed that this AT-rich sequence might result in an undesirable termination of S1 gene transcription by acting as a polyadenylation loop in P. pastoris. Based on the same meaning of the native codons, we changed A or T to G or C as much as possible at four sites with over 7 nucleotides of A + T repetition. As a consequence, this modification of S1 gene rendered its high-level expression in P. pastoris with the yield of 69 mg/l. The
expressed S1 was functionally demonstrated by its reaction with antisera to SARS-CoV and the receptor ACE2 (Figs. 3–5). These evidences support our assumption that high content of A + T in heterologous genes impedes the expression in *P. pastoris*.

One of the advantages for *P. pastoris* expression system is that the expressed protein can be glycosylated. However, the glycosylation patterns differ between yeast and mammal expression systems [19]. Although to add both O- and N-linked carbohydrates moieties to secreted proteins, *P. pastoris* glycosylated a heterologous protein by adding O-linked and N-linked high-mannose oligosaccharides. Some of foreign proteins appear to be hyperglycosylated with outer chains of 50–150 mannose residues in length [12]. The native S protein of SARS-CoV is a highly-glycosylated protein with high mannose and/or hybrid oligosaccharides, and our S1 protein fragment includes four N-glycosylation sites located in residues 269, 318, 330, and 357, respectively [20, 21]. The hyperglycosylation of S1 protein secreted by the recombinant *P. pastoris* was confirmed by PNGase F deglycosylation. PNGase F is an amidase, which cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid and complex oligosaccharides from N-linked glycoproteins. Therefore PNGase F hydrolyzes nearly all types of N-glycan chains from glycopeptides/proteins. After PNGase F deglycosylation, the high molecular mass (70 kDa) of S1 protein was degraded to the expected size of S1 (30 kDa). Unlike the SDS-PAGE result, we detected two bands on NC membrane when the antisera to SARS-CoV were used to probe the degraded S1 protein in Western blot analysis. Because the recognition site of PNGase F is located between the innermost GlcNAc and asparagine residues of high mannose, it would be possible that the other 40 kDa band was the oligosaccharide cleaved by PNGase F which could be recognized by antisera to SARS-CoV but not detected by protein stain Coomassie Brilliant Blue (Fig. 3).

In conclusion, this study successfully modified S1 gene of SARS-CoV by changing A or T in the native gene to G or C. This modification enabled the recombinant *P. pastoris* to express S1 protein with a high efficiency. The biological activity of expressed S1 protein was confirmed by reacting with the antisera and ACE2, the receptor of SARS-CoV. Based on the great concern on biosafety of live SARS-CoV, the high-level expression of functional S1 protein in *P. pastoris* is of significance to research on SARS-CoV pathogenesis and infection control.

**Acknowledgment** This work was supported by China National Basic Research Program (China “973” Program) (2003 CBS14122).

**References**

1. R.A. Fouchier, T. Kuiken, M. Schutten, G. van Amerongen, G.J. van Doornum, B.G. van den Hoogen, M. Peiris, W. Lim, K. Stohr, A.D. Osterhaus, Nature 423, 240 (2003)
2. P.A. Rota, M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Penaranda, B. Bankamp, K. Mahler, 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. Ksiazek, P.E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A.D. Osterhaus, C. Drosten, M.A. Pallansch, L.J. Anderson, W.J. Bellini, Science 300, 1394 (2003)
3. Chili-Fond Chon, Shuo Shen, Yee-Too Tan, B.C. Fielding, T.H.P. Tan, Jamlin Fu, Xu QiuRong, Seng Gee Lim, Wanjin Hong, J Virol Methods 123, 41 (2005)
4. S.A. Jeffers, S.M. Tussell, L. Gilim-Ross, E.M. Hemmila, J.E. Achenbach, G.B. Babcock, W.D. Thomas Jr., L.B. Thackray, M.D. Young, R.J. Mason, D.M. Ambrosino, D.E. Wentworth, J.C. Demartini, K.V. Holmes, Proc Natl Acad Sci USA 101, 15748 (2004)
5. W. Li, M.J. Moore, N. Vasileva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, H. Choe, M. Farzan, Nature 426, 450 (2003)
6. H.D. Song, C.C. Tu, G.W. Zhang, S.Y. Wang, K. Zheng, L.C. Lei, O.X. Chen, Y.W. Gao, H.Q. Zhou, H. Xiang, H.J. Zheng, S.W. Chen, F. Cheng, C.M. Pan, H. Xuan, S.J. Chen, H.M. Luo, D.H. Zhou, Y.F. Liu, J.F. He, P.Z. Qin, L.H. Li, Y.Q. Ren, W.J. Liang, Y.D. Yu, L. Anderson, M. Wang, R.H. Xu, X.W. Wu, H.Y. Zheng, J.D. Chen, G. Liang, Y. Gao, M. Liao, L. Fang, L.Y. Jiang, H. Li, F. Chen, B. Di, L.J. He, J.Y. Lin, S. Tong, X. Kong, L. Du, P. Hao, H. Tang, A. Bernini, X.J. Yu, O. Spiga, Z.M. Guo, H.Y. Pan, W.Z. He, J.C. Manuguerra, A. Fontanel, A. Danchin, N. Niccolai, Y.X. Li, C.I. Wu, G.P. Mao, Proc Natl Acad Sci USA 102, 2430 (2005)
7. Chun-Hua Hsu, Tzu-Ping Ko, Hui-Ming Yu, Tswen-Kei Tang, Shui-Tsong Chen, A.H.-J. Wang, Biochem Biophy Res Commun 324, 761 (2004)
8. P. Wang, J. Chen, A. Zheng, Y. Nie, X. Shi, W. Wang, G. Wang, M. Luo, H. Liu, L. Tan, X. Song, Z. Wang, X. Yin, X. Xu, W. Wang, T. Qing, M. Ding, H. Deng, Biochem Biophys Res Commun 315, 439 (2004)
9. S.K. Wong, W. Li, M.J. Moore, H. Choe, M. Farzan, J Biol Chem 279, 3197 (2004)
10. Y. He, H. Lu, P. Siddiqui, Y. Zhou, S. Jiang, J Immunol 174, 4908 (2005)
11. J. Sui, W. Li, A. Murakami, A. Tamin, L.J. Matthews, S.K. Wong, M.J. Moore, A.S. Tallarico, M. Olurinde, H. Choe, L.J. Anderson, W.J. Bellini, M. Farzan, W.A. Marasco, Proc Natl Acad Sci USA 101, 2536 (2004)
12. J.L. Cereghino, J.M. Cregg, FEMS Microbiol Rev 24, 45 (2000)
13. Mei-Ling Sun, Jin-Hua Piao, Chang-Cheng Yin, Gong-She Yang, Hua-Liang Huang, Hereditas (Beijing) 27, 623 (2005)
14. C.A. Scorer, R.G. Buckholz, J.J. Clare, M.A. Romanos, Gene 136, 111 (1993)
15. J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular cloning: a laboratory manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)
16. Liguo Yang, Pinghua Wei, Aizhen Guo, Shaochang Hu, Enzyme immunoassay technique. (Nanjing University Press, Nanjing, 1998)
17. Aizhen Guo, Chengping Lu, Chinese J Virol 16, 155 (2000)
18. M.A. Romanos, C.A. Scorer, T.T. Clare, Yeast 8, 423 (1992)
19. K. Sreekrishna, R.G. Brankamp, K.E. Kropp, D.T. Blakenship, J.T. Tsay, P.L. Smith, J.D. Wierschke, A. Subramaniam, L.A. Birkenberger, Gene 190, 55 (1997)
20. D.P. Han, H.G. Kim, Y.B. Kim, L.L. Poon, M.W. Cho, Virology 326, 140 (2004)
21. B. Nal, C. Chan, F. Kien, L. Siu, J. Tse, K. Chu, J. Kam, I. Staropoli, B. Crescenzo-Chaigne, N. Escriou, S. van der Werf, K.N. Yuen, R. Altmeyer, J Gen Virol 86, 1423 (2005)