

Dissection of SARS Coronavirus Spike Protein into Discrete Folded Fragments*

LI Shuang (李爽), CAI Zhen (蔡真), CHEN Yong (陈勇), LIN Zhanglin (林章凛)**

Laboratory of Directed Evolution, Department of Chemical Engineering, Tsinghua University,
Beijing 100084, China

Abstract: The spike protein of the severe acute respiratory syndrome coronavirus (SARS-CoV) mediates cell fusion by binding to target cell surface receptors. This paper reports a simple method for dissecting the viral protein and for searching for foldable fragments in a random but systematic manner. The method involves digestion by DNase I to generate a pool of short DNA segments, followed by an additional step of re-assembly of these segments to produce a library of DNA fragments with random ends but controllable lengths. To rapidly screen for discrete folded polypeptide fragments, the reassembled gene fragments were further cloned into a vector as N-terminal fusions to a folding reporter gene which was a variant of green fluorescent protein. Two foldable fragments were identified for the SARS-CoV spike protein, which coincide with various anti-SARS peptides derived from the heptad repeat (HR) region 2 of the spike protein. The method should be applicable to other viral proteins to isolate antigen or vaccine candidates, thus providing an alternative to the full-length proteins (subunits) or linear short peptides.

Key words: severe acute respiratory syndrome coronavirus (SARS-CoV); spike protein; dissection; foldable fragment; green fluorescent protein (GFP)

Introduction

Severe acute respiratory syndrome coronavirus (SARS-CoV), which is the causative agent of the atypical pneumonia, was first identified in the fall of 2002 to be a previously unknown member of the family of coronaviruses^[1]. The rapid transmission by means of aerosols and the high mortality rate (up to 10%) make SARS a potential global threat. An attractive approach to interfere with SARS disease progression focuses on one of the earliest infection processes by blocking the fusion process that mediates the

delivery of the viral genome into the host cell. The spike protein (S protein) is a 180- to 200-kDa type-I transmembrane glycoprotein that is responsible for the initiation and propagation of infection by interacting with a cellular receptor to induce cell-to-cell fusion. Binding of the S protein to a specific soluble or cell surface glycoprotein receptor induces global changes in the conformation of S protein that displays a previously hidden hydrophobic surface area which allows the virions to interact with the host cell membrane^[2,3]. Earlier attempted expressions of this S protein failed to obtain soluble full-length polypeptides in *Escherichia coli* (*E. coli*). Subsequent work aimed at identifying smaller but folded SARS-CoV spike fragments for use as possible antigen or vaccine candidates. The approach involved digestion and reassembly of the target gene to generate a pool with smaller DNA fragments of random ends but controllable lengths which were

Received: 2005-12-31; revised: 2006-02-21

* Supported by the Tsinghua University SARS Special Fund and the National Key Basic Research and Development (973) Program of China (No. 2003CB716002)

** To whom correspondence should be addressed.

E-mail: zhanglinlin@tsinghua.edu.cn; Tel: 86-10-62794403

screened for foldable fragments using a green fluorescent protein as a folding reporter^[4,5]. Two foldable fragments were identified, which coincide with various SARS peptides reported to have SARS neutralization activity^[6-8]. This dissection approach has the potential to be a generally applicable tool for producing foldable fragments of viral surface proteins that may provide discontinuous epitopes. These fragments should be easier to express in *E. coli* or other recombinant hosts.

1 Materials and Methods

1.1 pET30a-linker-GFP construction

The green fluorescent protein (GFP) gene was amplified from an in-house GFP-containing vector pET30a-*hydA* (Wang and Lin, unpublished result), which was in turn constructed from pQB-2^[9], then ligated into the pET30a(+) (Novagen) to yield the pET30a-linker-GFP.

1.2 Fragment library construction

The SARS-CoV spike gene was obtained from the Huada Beijing Genomics Institute. Fragmentation and re-assembly of the target gene were performed as described by Lorimer and Pastan^[10]. The reassembled DNA sample was then purified and phosphorylated with T4 polynucleotide kinase at 37°C for 30 min. The backbone vector pET30a-linker-GFP was digested with *EcoR* I, blunt-ended with T4 DNA polymerase in the presence of 0.1 mmol/L each dNTP, and purified with a QIAgen[®] gel purification kit to remove residual enzyme activity. The linearized and blunt-ended vector was then dephosphorylated with shrimp alkaline phosphatase (SAP) followed by heat denaturation to deactivate the enzyme. The gene fragments and the backbone vector were ligated at 12°C overnight in the presence of 5% PEG8000 and then transformed into *E. coli* BL21(DE3) (Novagen) competent cells by electroporation.

1.3 Screening of fragments

Transformed *E. coli* BL21(DE3) cells were plated on Luria-Bertani (LB) medium supplemented with 50 µg/mL kanamycin and grown overnight at 37°C, then grown further on a bench for about 20 h. The fluorescent colonies were picked and tested with colony PCR by using primers flanking the fragment inserts, and sequenced. No isopropylthio-β-D-galactoside (IPTG)

was used in these experiments, as it would inhibit the formation of fluorescent colonies.

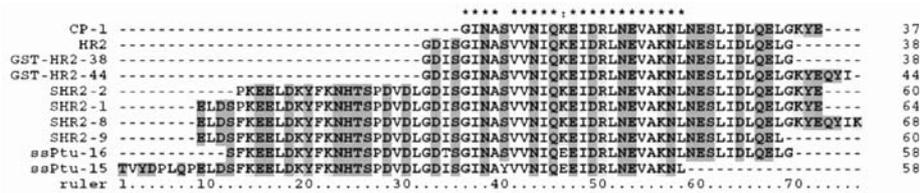
1.4 Expression analysis of fusion proteins

Saturated overnight cultures were diluted 100-fold into LB medium containing 50 µg/mL kanamycin and grown at 37°C for about 2 h to reach an optical density at 600 nm (OD₆₀₀) of 0.5-0.6. Protein expression was initiated with 0.2 mmol/L of IPTG, and continued for 4 h at 23°C. Cells were then collected and lysed for soluble protein extraction. The supernatant fractions (soluble protein) and cell pellets (insoluble protein) were resolved by SDS-PAGE using a 12% acrylamide gel.

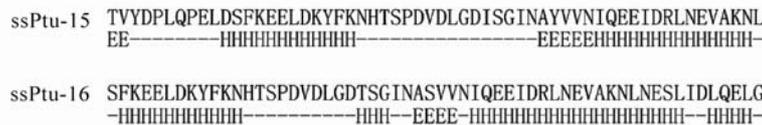
2 Results and Discussion

This work sought to identify smaller but folded SARS-CoV spike fragments for use as possible antigen or vaccine candidates. Compared with linear short peptides derived from the protein, folded fragments may be advantageous as they have the potential to provide discontinuous epitopes. The SARS-CoV spike gene was digested by DNase I to generate a pool of short DNA segments, followed by an additional step of reassembly of these segments to produce a library of DNA fragments with random ends^[10,11]. This is in part analogous to the DNA shuffling protocol^[12,13], but the purpose here is not to produce full-length hybrids from a group of different parental genes, but to generate various smaller DNA fragments from a single template gene. The reassembly step following the DNase I treatment is necessary to prepare a large number of DNA sequences with controlled lengths, which was achieved by tailoring the number of PCR cycles used in the reassembly (see Methods). To screen for discrete folded polypeptide fragments, the reassembled gene fragments were further cloned into a vector as N-terminal fusions to a folding reporter gene which was a variant of the green fluorescent protein that exhibits strong fluorescence upon UV excitation^[9]. GFP has been shown to be an effective indicator for the foldability of the upstream polypeptide partner^[4,5]. The vector construction is shown in Fig. 1.

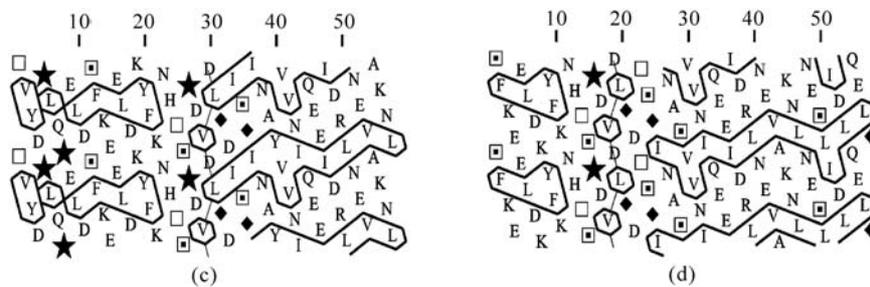
Among about 4300 clones screened, 230 clones were found to be fluorescent (see Fig. 2a). These clones were then subjected to rapid colony PCR



(a) CLUSTALW alignment of ssPtu-15 and ssPtu-16 with HR2 derived peptides which interfere with SARS-CoV S-mediated fusion to host cells: peptide CP-1^[7], peptides HR2, GST-HR2-38, GST-HR2-44^[8], peptides SHR2-1, SHR2-2, SHR2-8, and SHR2-9^[6].



(b) Secondary structure prediction for ssPtu-15 and ssPtu-16 by 3D-Jury (<http://bioinfo.pl/Meta/>)^[19]. E, β -strand; H, α -helix.



(c) and (d) Hydrophobic clusters analysis (HCA) plot for ssPtu-15 and ssPtu-16 drawn using DRAWHCA (<http://bioserv.rpbs.jussieu.fr>)^[17]. Protein sequences are displayed on a duplicated helix using one-letter codes for the amino acids except for prolines (★), glycines (◆), threonine (□), and serine (◻). Hydrophobic residues are automatically contoured.

Fig. 3 Sequence analysis

Several studies^[6-8] have reported that SARS-CoV S-mediated fusion can be inhibited by HR2 but not HR1-derived peptides, most likely by interfering with the six-helix bundle formation, a process essential to drive the membrane fusion reaction and to initiate infection^[14]. For the majority of these peptides, micromolar concentrations were required for efficient inhibition of the viral infection, indicating that although these peptides are effective, further optimization is required to achieve efficient inhibition of SARS-CoV in infected individuals. Given the high similarity of ssPtu-15 and ssPtu-16 with these peptides derived from the HR2 region^[6-8], ssPtu-15 and ssPtu-16 may both have potential as therapeutic agents for the direct inhibition of SARS-CoV cell entry, as an anti-SARS vaccine, and as a high throughput assay for screening for small molecule inhibitors of SARS envelope-mediated cell fusion.

In summary, the dissection approach described in this study has the potential to produce foldable fragments of viral surface proteins that may be useful for the design of antiviral compounds and provide

alternative antigen or vaccine candidates. The method is target protein independent and thus can be applied to various viral proteins. The process is also simple and rapid. The method should be applicable for dissecting and understanding other non-viral proteins, for example, to identify smaller polypeptide units that are structurally, functionally, or evolutionally relevant.

Acknowledgements

The authors thank Dr. K. Kohno and Dr. X. Xing for the GFP gene and the Huada Beijing Genomics Institute and Dr. Y. Ruo for the SARS-CoV spike gene.

References

- [1] Rota P A, Oberste M S, Monroe S S, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003, **300**(5624): 1394-1399.
- [2] Melikyan G B, Markosyan R M, Hemmati H, et al. Evidence that the transition of HIV-1 gp41 into a six-helix

- bundle, not the bundle configuration, induces membrane fusion. *J. Cell Biol.*, 2000, **151**(2): 413-423.
- [3] Holmes K V, Zelus B D, Schickli J H, et al. Receptor specificity and receptor-induced conformational changes in mouse hepatitis virus spike glycoprotein. *Nidoviruses (Coronaviruses and Arteriviruses)*, 2001, **494**: 173-181.
- [4] Waldo G S, Standish B M, Berendzen J, et al. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.*, 1999, **17**(7): 691-695.
- [5] Waldo G S. Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.*, 2003, **7**(1): 33-38.
- [6] Bosch B J, Martina B E E, van der Zee R, et al. Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**(22): 8455-8460.
- [7] Liu S W, Xiao G F, Chen Y B, et al. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: Implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet*, 2004, **363**(9413): 938-947.
- [8] Zhu J Q, Xiao G F, Xu Y H, et al. Following the rule: Formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem. Biophys. Res. Commun.*, 2004, **319**(1): 283-288.
- [9] Kimata Y, Iwaki M, Lim C R, et al. A novel mutation which enhances the fluorescence of green fluorescent protein at high temperatures. *Biochem. Biophys. Res. Commun.*, 1997, **232**(1): 69-73.
- [10] Lorimer I A J, Pastan I. Random recombination of antibody single-chain Fv sequences after fragmentation with DNase I in the presence of Mn²⁺. *Nucleic Acids Res.*, 1995, **23**(15): 3067-3068.
- [11] Prodromou C, Pearl L. Recursive PCR: A novel technique for total gene synthesis. *Protein Eng.*, 1992, **5**(8): 827-829.
- [12] Stemmer W P C. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, 1994, **370**(6488): 389-391.
- [13] Joo H, Lin Z L, Arnold F H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature*, 1999, **399**(6737): 670-673.
- [14] Eckert D M, Kim P S. Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem.*, 2001, **70**: 777-810.
- [15] Skehel J J, Wiley D C. Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell*, 1998, **95**(7): 871-874.
- [16] Tripet B, Howard M W, Jobling M, et al. Structural characterization of the SARS-coronavirus spike (S) fusion protein core. *J. Biol. Chem.*, 2004, **279**(20): 20 836-20 849.
- [17] Gaboriaud C, Bissery V, Benchetrit T, et al. Hydrophobic cluster analysis: An efficient new way to compare and analyze amino acid sequences. *FEBS Lett.*, 1987, **224**(1): 149-155.
- [18] Harbury P, Zhang T, Kim P, et al. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, 1993, **262**(5138): 1401-1407.
- [19] Ginalski K, Elofsson A, Fischer D, et al. 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics*, 2003, **19**(8): 1015-1018.