Chapter 18

Identification of Viral Peptide Fragments for Vaccine Development

Zhanglin Lin, Shuang Li, and Yong Chen

Summary

We report a simple method for identifying foldable viral surface protein fragments in a random but systematic manner. The method involves digestion and reassembly of a target gene to generate a pool of smaller DNA fragments with random ends but controllable lengths, followed by screening for foldable fragments using green fluorescent protein (GFP) as a folding reporter. The surface glycoproteins of SARS-CoV and HIV-1 were used as model proteins. Two foldable fragments for SARS-CoV spike protein were identified, which coincide with various anti-SARS-CoV peptides. A similar treatment of the HIV-1 gp120 yielded a number of fragments that are associated with the critical CD4 binding site, or the partially buried CCR5 binding site of the protein. The random dissection approach described here should be applicable to other viral proteins for isolating soluble viral surface protein fragments, and may provide alternatives to the full-length proteins (subunits) or linear short peptides in search for antigen or vaccine candidates.

Key words: Vaccine development, Peptide fragment, Surface glycoprotein, SARS, CoV, HIV-1, gp120.

1. Introduction

The entry of viruses into target cells depends on sequential interactions of the viral surface proteins with the cellular receptors, and thus these surface proteins are important for developing antiviral treatments *(1)*. However, it is often difficult to express viral surface proteins in facile recombinant hosts such as *Escherichia coli* (*E. coli*), and this hinders antibody or vaccine development. We set out to find a way to identify smaller but soluble viral surface protein fragments for use in screening for possible antigen or vaccine candidates. Compared with linear peptides derived from

the surface proteins, soluble fragments may be advantageous as they have the potential to provide discontinuous epitopes. Additionally, it has been suggested that the binding of viral surface proteins to cell receptors may expose sites in the viral proteins otherwise inaccessible *(2, 3)*. In particular, the binding of HIV gp120 to CD4 seems to expose a site that is critical for subsequent binding to the CCR receptor, but which is partially buried in the resting state of the HIV surface protein. This has been implicated as a possible mechanism adopted by the virulent virus to evade the immune surveillance *(2, 3)*. We reasoned that by producing discrete foldable fragments which are taken out of the context of the full-length protein, some of these hidden critical sites might become exposed in the fragments thus obtained. Consequently, these discrete foldable polypeptide fragments might facilitate screening for more effective vaccines or antigen candidates.

Searching for soluble fragments of a polypeptide chain may be achieved by genetic cleavage at predetermined sites based on detailed analysis of the target structure, or by the more traditional protease digestion method *(4–7)*. However, the former approach is limited by the generally poor understanding of the structure–function relationship of proteins, while for the latter only a small number of exposed surface sites are accessible to proteases. A random method that in theory can thoroughly and individually analyze all possible folding fragments would be more desirable. To this end, we formulated a dissection scheme that is capable of searching for soluble fragments of a given target in a random but systematic manner, and which is independent of protein function. The solubility of the dissected protein fragments was accessed through expression analysis of these fragments as N-terminal fusions to green fluorescent protein (GFP), which has been shown to be a good indicator for the "foldability" of the upstream polypeptide partner *(8, 9)*. This method using GFP as folding reporter is generally applicable because it does not require structural or functional information on the target proteins. Results from dissection of the severe acute respiratory syndrome coronavirus (SARS-CoV) spike protein and HIV-1 gp120 are presented.

2. Materials

2.1. Reagents, Buffers, and Solutions

1. PEG 8000 (Ameresco, Farmingham, MA).
2. dNTPs (TaKaRa, Dalian, China).
3. LB medium.
4. Kanamycin (BBI, Ont., Canada).
5. IPTG (isopropylthio-β-D-galactoside) (Ameresco).
6. Lysozyme (Sigma).

7. Coomassie brilliant blue R250.

8. Lysis buffer: 50 mM Tris-HCl, pH 7.2, 50 mM NaCl, and 5% glycerol.

9. DNase I random digestion buffer: 50 mM Tris-HCl, pH 7.5, 10 mM $MnCl_2$.

10. Fragment reassembly buffer: 10 mM Tris-HCl, pH 8.3, 2 mM $MgCl_2$, 50 mM KCl, 0.2 mM each dNTP.

11. Loading buffer: 50 mM Tris-HCl, pH 6.8, 0.1% bromophenol blue, 2% SDS, 1% 2-mercaptoethanol, and 10% glycerol.

12. Prestained protein markers (New England Biolabs, Beverly, MA).

2.2. Strain and Plasmid

1. *E. coli* strains BL21(DE3) (Novagen, Madison, WI).

2. pET30a(+) expression system (Novagen).

3. Oligonucleotide primers (TaKaRa).

2.3. Enzymes

1. Restriction enzymes (New England Biolabs).

2. Taq polymerase (TaKaRa).

3. Deep Vent® polymerase (New England Biolabs).

4. T4 DNA ligase (New England Biolabs).

5. DNase I (Worthington, Lakewood, NJ).

6. T4 DNA polymerase (New England Biolabs).

7. T4 polynucleotide kinase (New England Biolabs).

8. Shrimp alkaline phosphatase (Promega, Madison, WI).

2.4. Kits and Apparatus

1. QIAquick Gel Extraction Kit, QIAquick PCR Purification Kit, QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA).

2. Thermocycler (MJ research PTC-225, Miami, FL).

3. Electroporator (Eppendorf, Hamburg, Germany).

4. Agarose gel equipment.

5. Sodium dodecyl sulfate-polyacrymide gel electrophoresis (SDS-PAGE) equipment.

6. Microcon® YM-30 (Millipore, Billerica, MA).

3. Methods

The methods described below outline (1) the construction of the expression plasmid, (2) viral protein fragment library construction, (3) screening of fragments, and (4) expression analysis of viral protein fragments.

3.1. Expression Plasmid

The construction of the expression plasmid pET30-linker-GFP for viral gene fragments is described in the **Subheadings 3.1.1–3.1.2**. This includes (a) the description of the expression vector of pET30a(+), (b) the description of the GFP gene, and (c) the cloning strategy.

3.1.1. pET30a(+) Expression Vector and cDNA of GFP

1. Use the T7 promoter-driven system originally developed by Studier and colleagues *(10–12)*, pET30a(+) (*see* **Fig. 1a**). It is a powerful vector for the cloning and expression of recombinant proteins in *E. coli*. Target genes can be inserted into the multiple cloning region and placed under the control of strong bacteriophage T7 transcription and translation signals, and expression is induced by isopropylthio-β-D-galactoside (IPTG). The pET30a(+) expression vector carries an N-terminal His•Tag configuration plus an optional C-terminal His•Tag sequence which can be used to assay expression levels and purify proteins. The pET30a(+) expression vector also contains a kanamycin resistance gene.
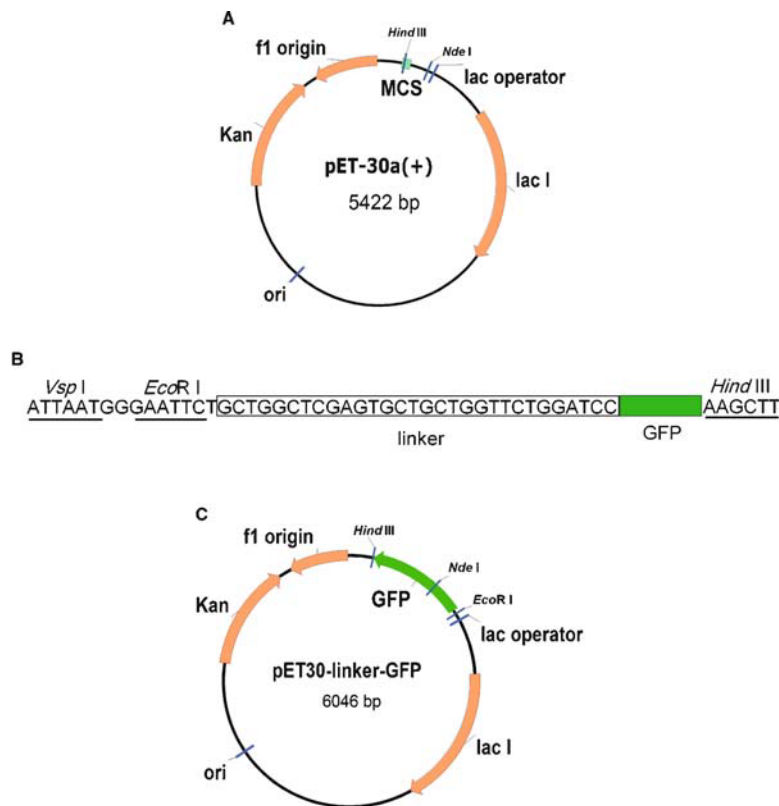

Fig. 1a. Schematic drawing of pET30a(+) expression vector, adapted from Navogen (Madison, WI). b Expression feature contained in pET30-linker-GFP vector, which otherwise is identical to pET30a(+). It contains a linker sequence AGSSAAGSGS (*boxed*) upstream of the GFP gene and an internal *Eco*RI site (*underlined*) used for insertion of gene fragments. c Schematic drawing of pET30-linker-GFP vector.

2. Subclone the GFP gene from a commercial source. We used an in-house GFP-containing vector pET30a-*hydA*, which was constructed from pQB-2 *(13)*. This GFP variant carries 11 extra amino acid residues at the C-terminus which help emit a stronger fluorescent signal than the wild-type GFP *(13)*.

3. Amplify the GFP gene using *LA* Taq polymerase (TaKaRa) and the following primers (forward and reverse, respectively):

5′-CATCGTT**ATTAAT**GGG**GAATTC**T<u>GCTGGCTC-GAGTGCTGCTGGT</u>-3′, and 5′-TAG**AAGCTT**AGCTAAT-TCA-GCTTGGCTGC-3′.

The restriction endonuclease sites in these primers are *Vsp* I, *Eco*RI, and *Hin*dIII (shown in bold), respectively, and the partial linker sequence is underlined) (**Fig. 1b**).

3.1.2. pET30-Linker-GFP Construction

1. Doubly digest the amplified GFP gene with *Vsp*I and *Hin*dIII.

2. Ligate the insert into the pET30a(+) plasmid which has been doubly digested with *Nde*I and *Hin*dIII to yield pET30-linker-GFP (*see* **Note 1**).

3. Transform the expression vector into *E. coli* BL21(DE3) cells by standard methods *(14)*.

4. Plate the transformed *E. coli* BL21(DE3) cells on LB plates containing kanamycin (50 µg/mL) and incubate overnight at 37 ºC.

5. Select single colonies and grown overnight in LB medium with kanamycin. This can be done in 96-well plates.

6. Isolate the plasmid DNA using a QIAprep Spin Miniprep Kit (QIAGEN), and verify sequence of the region flanking GFP (**Fig. 1c**).

3.2. Fragment Library Construction

Described below are the steps for the construction of a viral surface protein fragment library (**Fig. 2**).

3.2.1. cDNA

SARS Spike Protein

1. Amplify the gene of interest for vaccine development by PCR using Taq polymerase. The SARS-CoV spike gene (GenBank No. AY278488) we used was amplified by PCR reaction from a cDNA sample kindly donated by Huada Beijing Genomics Institute containing the full length cDNA for SARS-CoV spike protein.

2. The primers used were (forward and reverse, respectively): 5'-CGGAATTCCATATGTTTATTTTCTTATTATTTC-3′ and 5′-CCGGATCCTTAGTGGTGGTGGTGGTGGTGT-GTGTAATGTAATTTGACACC-3′.

HIV-1 gp120

If a vector containing the gene is not available but its sequence is known, it can be synthesized according to DNAworks *(15, 16)* and confirmed by sequencing (*see* **Note 2**). We used this strategy to obtain the HIV-1 gp120 gene.
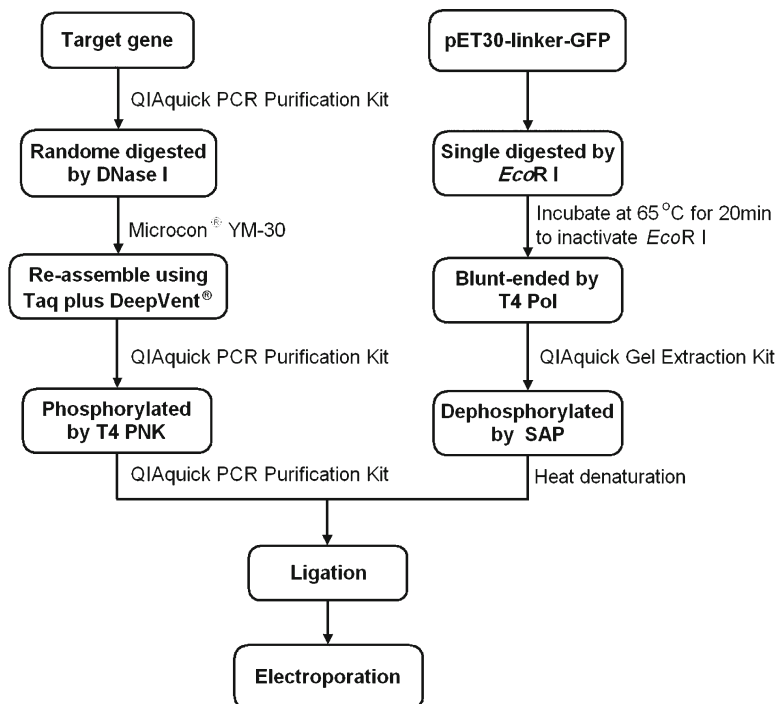
Fig. 2. Schematic drawing of fragment library construction.

3.2.2. Random Dissection
of Target Genes

Fragmentation of both of the target genes (i.e., SARS-CoV spike
and HIV-1 gp120) is performed by digestion with a nonselective
DNase I (see **Note 3**) (17).

1. Purify 2–4 μg of the target gene using a QIAquick PCR Purification Kit.

2. Incubate the purified DNA in 50 μL DNase I random digestion buffer for 10 min at 15 °C.

3. Add 0.075 U of DNase I and incubate at 15 °C for 4 min.

4. Add 10 μL of 0.5 M EDTA to terminate the DNA digestion.

5. A pool of short DNA segments ranging from ~50 to 100 bp should be prepared. Verify this by running a 1–4% agarose gel with a DNA ladder containing a 100-bp marker (**Fig. 3a**).

3.2.3. Reassembly of Gene
Fragments

1. Purify the randomly digested gene fragments by passing the reaction mixture through a Microcon® YM-30 column (see **Note 4**).

2. Resuspend the product in the fragment reassembly buffer at a concentration of 1 ng/μL.

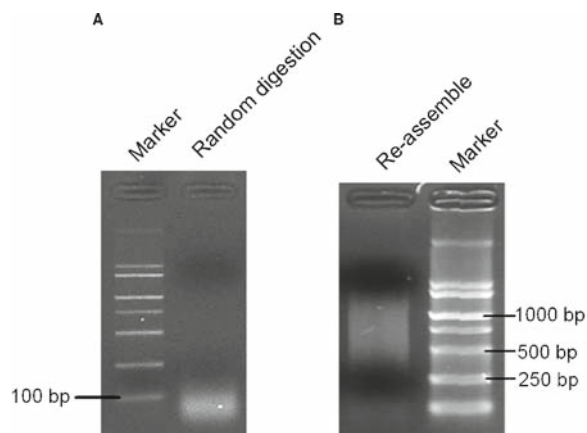3. Add 2.5 U/100 μL of both Taq polymerase and Deep Vent® polymerase.

Fig. 3. Fragment library construction for SARS-CoV spike. **a** Fragmentation of target gene digested by DNase I. **b** Reassembled gene fragments.

4. Using an MJ Research PTC-225 thermocycler or comparable instrument, run a PCR program (*18*) consisting of 10–20 cycles of 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min + 5 s/cycle (*see* **Note 5**).

5. This should generate gene fragments largely in the range of 200–1,200 bp (**Fig. 3b**). Verify this by agarose gel electrophoresis with a ladder spanning 100–2,000 bp.

3.2.4. Cloning

1. Purify the reassembled DNA sample using a QIAquick PCR Purification Kit.

2. Phosphorylate the DNA with T4 polynucleotide kinase (T4 PNK) at a ratio of 5 U/μg DNA by incubating at 37 °C for 30 min.

3. Digest the backbone vector (pET30-linker-GFP) with *Eco*RI.

4. Make blunt-ended vector DNA using T4 DNA polymerase in the presence of 0.1 mM of each dNTP (*see* **Note 6**).

5. Purify the DNA with a QIAquick Gel Extraction Kit to remove residual enzyme activity.

6. Dephosphorylate the linearized and blunt-ended vector twice with shrimp alkaline phosphatase (SAP) at a ratio of 10 U/μg DNA by incubating at 37 °C for 45 min in the supplied SAP buffer.

7. Incubate at 70 °C for 20 min to inactivate the alkaline phosphatase enzyme (*see* **Note 7**).

8. Ligate the gene fragments to the backbone vector at 12 °C overnight in the presence of 5% PEG 8000.

9. Transform *E. coli* BL21(DE3) competent cells by electroporation.

10. Plate transformed *E. coli* BL21(DE3) cells on LB agar supplemented with 50 μg/mL kanamycin and grow overnight at 37 ºC. Continue growing cells at ambient temperature on the bench for about 20 h (*see* **Note 8**).

11. No IPTG is added in these experiments, as it would inhibit the formation of fluorescent colonies (*see* **Note 9**).

3.3. Screening of Soluble Viral Protein Fragments

3.3.1. Screening Scheme

1. The transformed fluorescent colonies are picked under a UV lamp and tested with standard colony PCR using primers flanking the fragment inserts, and sequenced. Alternatively, the colonies are left on bench at room temperature for additional 24 h to allow for color development, and then fluorescent colonies are picked directly.

2. The primers we used in the colony PCR were (forward and reverse, respectively): 5′-TAAGAAGGAGATATACATAATG-3′ and 5′-AGAACCAGCAGCACTCGAGCCA-3′.

    These primers allowed determination of the size of the inserted fragment.

3.3.2. Primary Analysis of Screened Fragments

SARS-CoV Spike Protein

1. Screen as many clones as necessary/possible. For example, from ~4,300 clones screened, only 230 clones were found to be fluorescent (*see* **Fig. 4a**).

2. Most of the fluorescent clones will contain empty vectors or vectors with fragments smaller than 100 base pairs (bp) as judged by colony PCR and sequencing.
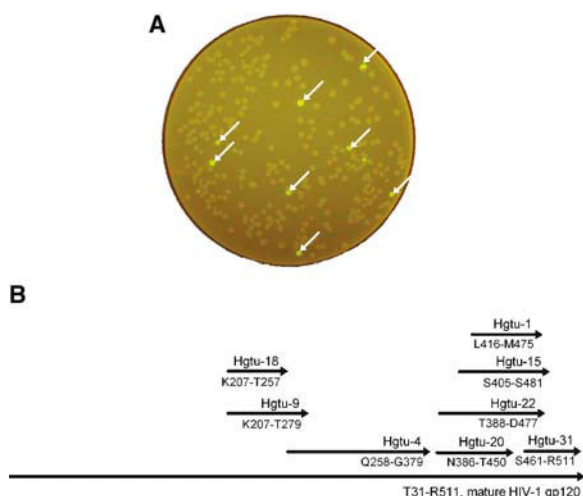

Fig. 4a. Colonies obtained from inserting and expression of SARS-CoV spike gene fragments in pET30-linker-GFP using *E. coli* BL21(DE3) as the host. The fluorescent clones were indicated by *arrows*. b Fragments of HIV-1 gp120 protein as deduced from the sequences of the inserts contained in fluorescent GFP-fusion clones. The numbering of the residues is that used in the X-ray structure of a truncated HIV-1 gp120 *(2)*. Only fragments larger than 50 deduced amino acid residues are shown. The start and stop positions of the deduced amino acid sequence for each fragment are indicated beneath the respective *arrow*.

3. In addition, ~20 of the 230 fluorescent colonies were found to contain vectors with inserts in the reverse orientation or not in frame, so sequencing all of the colonies with appropriately sized fragments is absolutely necessary.

4. Screen the peptides by SDS-PAGE.

5. Judged by SDS-PAGE results, many of the peptides encoded by these gene inserts will be degraded in the corresponding fusion proteins (data are not shown).

6. We were able to locate two inserts larger than 150 nucleotides (50 amino acid residues). These were labeled ssPtu-15 (residues 1,118–1,175 of the original protein) and ssPtu-16 (residues 1,129–1,186).

HIV-1 gp120

Out of 2,800 clones that we screened, 115 were found to be fluorescent, from which eight fragments containing greater than 50 deduced amino acid residues were isolated (**Fig. 4b**).

3.4. Expression Analysis of Viral Protein Fragments

3.4.1. Preliminary Structural Examination

SARS-CoV Spike

Several studies *(19–21)* have reported that SARS-CoV S-mediated fusion can be inhibited by heptad repeat region 2 (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 *(1)*.

1. By alignment with the HR2 region using Clustal W *(22)*, we later found that the SARS fragment ssPtu-15 overlaps with the HR2 (residues 1,147–1,185) of SARS-CoV spike protein *(23)*, while fragment ssPtu-16 contains the whole SARS HR2 (**Fig. 5**).

2. Given the high similarity of ssPtu-15 and ssPtu-16 with these peptides derived from the HR2 region *(19–21)*, ssPtu-15 and ssPtu-16 may share potential as therapeutic agents for the direct inhibition of SARS-CoV cell entry, as anti-SARS vaccines, and high throughput assay reagents for screening for small molecule inhibitors of SARS envelope-mediated cell fusion.
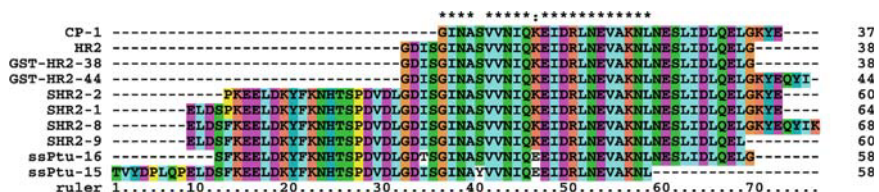

Fig. 5. Clustal W *(22)* alignment of ssPtu-15 and ssPtu-16 with HR2-derived peptides which interfere SARS-CoV S-mediated fusion to host cells: peptide CP-1 *(20)*, peptides HR2, GST-HR2-38, GST-HR2-44 *(21)*, peptides SHR2-1, SHR2-2, SHR2-8, SHR2-9 *(19)*.

Table 1
Deduced HIV-1 gp120 fragments and their locations in the protein structure

Fragment number	Number of residues	Location in the protein structure[a]
Hgtu-1	60	β19 – loop(β24~α5)
Hgtu-4	122	loopB – β16
Hgtu-9	72	loop (β3~β4) – β9
Hgtu-15	77	loopV4 – α5
Hgtu-18	51	loop (β3~β4) – loopB
Hgtu-20	65	β17 – β23
Hgtu-22	90	α4 – α5
Hgtu-31	51	loopV5 – C-terminal

[a]According to the assignment of structural elements used in the X-ray structure of HIV-1 gp120 *(2)*

HIV-1 gp120

The eight fragments were mapped onto the HIV-1 gp120 structure as shown in **Table 1**. A preliminary structural examination suggested that most of these fragments are associated with the proposed binding sites for CD4 and/or the chemokine receptor CCR5 *(2)*.

3.4.2. Fragment Expression Analysis

Induction and Protein Extraction

1. Dilute saturated overnight cultures 100-fold into fresh LB media containing 50 μg/mL kanamycin and grow at 37 °C for about 2 h to reach an optical density at 600 nm (OD_{600}) of 0.5–0.6.

2. Initiate protein expression by adding 0.2 mM IPTG and culture for 24 h at 23 °C.

3. Collect about 20 OD_{600} of the cells by centrifugation (3,500 rpm, 10 min) and resuspend in 1 mL of lysis buffer.

4. Sonicate for 66 pulses of 3 s each with a 3-s interval in an ice-water bath.

5. Supplement the lysate with 0.2 mg/mL of lysozyme and shake gently for 1 h at room temperature.

6. Centrifuge at 16,100 rpm for 5 min to separate soluble protein from cell debris and inclusion bodies.

Fragment-GFP Fusions Characterized by SDS-PAGE

1. Collect the supernatants, and resuspend the pellets in 1 mL of 2× SDS polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer. Likewise prepare supernatant samples for SDS-PAGE.

2. Resolve protein samples using a 12% acrylamide gel, and stain with coomassie brilliant blue dye. Typically, 10 μL aliquots were loaded.

3. Analysis by SDS-PAGE of fragment-GFP fusions found from both of SARS spike protein fragments (ssPtu-15, ssPtu-16) and three of the eight HIV-1 gp120 fragments (Hgtu-1, Hgtu-4, and Hgtu-15) to be partially soluble (*see* **Fig. 6**). Under the conditions we have tested so far, no significant soluble expression was seen for the other five HIV-1 gp120 fragments as determined by coomassie brilliant blue staining, while inclusion bodies were observed. Thus, Hgtu-1, Hgtu-4, and Hgtu-15 might be more useful as possible antigen and vaccine candidates.
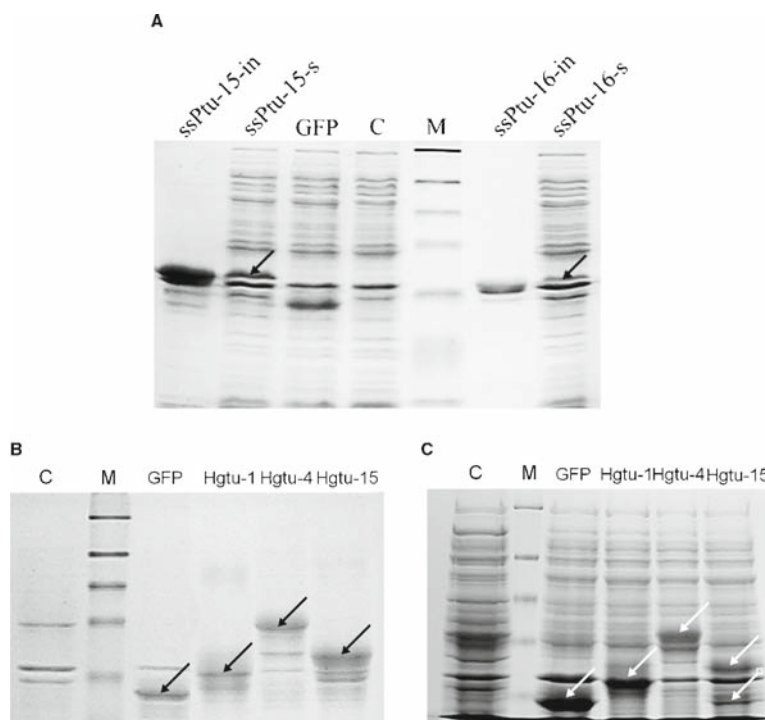

Fig. 6. SDS-PAGE analysis of viral protein fragment-GFP fusions, with *E. coli* BL21(DE3) as control. a SARS-CoV spike protein fragment-GFP fusions. "s" indicates supernatants of lysates, and "*in*" denotes insoluble pellets of the lysates. b Supernatants of the lysates for selected HIV-1 gp120 fragment-GFP fusions. c Pellets of the lysates for selected HIV-1 gp120 fragment-GFP fusions. Calculated molecular weights for GFP, ssPtu-15, ssPtu-16, Hgtu-1, Hgtu-4, and Hgtu-15 were 29.9, 36.3, 36.3, 36.5, 43.5, and 38.6 kDa, respectively. Corresponding band positions are indicated by *arrows*. *C*: control; *M*: protein marker, broad range (NEB), whose bands were 175, 83, 62, 48, 33, 25, and 17 kDa, respectively.

Notes

1. The 5′ cohesive end of the PCR product containing the GFP gene and digested with *Vsp*I is compatible with the 3′ end of pET30a(+) plasmid digested with *Nde*I. The reason for the use of *Vsp*I is that there is an *Nde*I restriction site in the GFP gene.

2. By inputting the gene sequence of HIV-1 HBc2 gp120 (1,446 bp) into the free software DNAworks (http://mcl1.ncifcrf.gov/dnaworks/dnaworks2.html), 66 oligo sequences were returned, and synthesized accordingly *(15)*. These oligos were assembled to gp120 by the standard overlapping procedure *(24)*. Then two flanking primers (5′-ATGACC-GAAAAACTGTGGGTGA-3′ and 5′-AGCGCTTCTCAC-GTTGAACAACACG-3′) were used to amplify the full-length of gp120 by touch-down PCR.

3. In the random digestion step, the use of DNase I in the presence of MnCl$_2$ is critical as this protocol will generate DNA fragments of relatively uniform sizes, which facilitates the reassembly step (**17**, also *see* **Note 5**). Other digestion methods often lead to smears of fragments that are difficult to purify and use. Do not use the buffer supplied with the DNase I enzyme, which contains MgCl$_2$ and would lead to smears of digestion products. Additionally, the amounts of the target DNA, DNase I, and the digestion time should be controlled carefully as indicated above. Otherwise, the DNA would be overdigested or underdigested and could not produce appropriately sized fragments.

4. When the random digestion by DNase I is used for some genes, peculiarly a small amount of the full-length gene has been observed even after prolonged incubation. In this case, the digestion mixture should be passed through a column with an appropriate molecular weight cut-off to filter out the full-length gene, followed by a Microcon® YM-30 column to accomplish the buffer exchange.

5. The manipulation in **Subheadings 3.2.2** and **3.2.3** is in part analogous to the DNA shuffling protocol *(24, 25)*, but unlike the latter, the purpose here is not to produce full-length hybrids from a group of different parental genes, but to generate smaller and different DNA fragments for a single template gene. This reassembly step following DNase I treatment is necessary in order to prepare a large amount of DNA sequences with controlled lengths, which is achieved by tailoring the PCR cycle number used in the reassembly.

6. In our work, several enzymes, that is, T4 DNA polymerase, S1 Nuclease, Mung Bean Nuclease, and Deep Vent® polymerase,

were tested for blunt-ending gene fragments. T4 DNA polymerase yielded the best results.

7. In order to improve the final yield of backbone vector, the manipulation steps should be as few as possible. The heat-sensitive shrimp alkaline phosphatase was used to reduce one purification procedure after dephosphorylation.

8. In order to improve the percentage of positive clones containing larger inserts (inserts of more than 50 amino acids), three principal improvements were made. First, gene fragments resulting from reassembly were phosphorylated using T4 polynucleotide kinase. Second, T4 polymerase was denatured with a QIAquick Gel Extraction Kit upon the completion of the blunt-ending reaction to prevent residual T4 DNA polymerase activity from degrading DNA in later steps. Third, a second step of the dephosphorylation by heat-liable SAP was incorporated to improve dephosphorylation efficiency.

9. We found that the basal expression level often observed for the pET vectors was sufficient for the expression of gene fragment-GFP fusion. Low IPTG concentrations (lower than 0.1 mM) did not enhance green fluorescent colonies, while higher concentrations only inhibited fluorescent colonies or generated abnormal hallow colonies.

Acknowledgments

References

1. Eckert, D. M., and Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition. *Annu. Rev. Biochem.* 70, 777–810.

2. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659.

3. Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A., and Sodroski, J. (1998) A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* 280, 1949–1953.

4. Collinet, B., Herve, M., Pecorari, F., Minard, P., Eder, O., and Desmadril, M. (2000) Functionally accepted insertions of proteins

within protein domains. *J. Biol. Chem.* 275, 17428–17433.

5. Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a (βα)8-barrel enzyme into two folded halves. *Nat. Struct. Biol.* 8, 32–36.

6. Smith, V. F., and Matthews, C. R. (2001) Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from *E. coli*: fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci.* 10, 116–128.

7. Soberon, X., Fuentes-Gallego, P., and Saab-Rincon, G. (2004) *In vivo* fragment complementation of a (β/α)8 barrel protein: generation of variability by recombination. *FEBS Lett.* 560, 167–172.

8. Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695.

9. Waldo, G. S. (2003) Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* 7, 33–38.

10. Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130.

11. Rosenberg, A. H., Lade, B. N., Dao-shan, C., Lin, S.-W., Dunn, J. J., and Studier, F.W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* 56, 125–135.

12. Studier, F. W., Rosenberg, A. H., Dunn, J. J., Dubendorff, J. W., and David,V. G.(1990) Use of T7 RNA polymerase to direct expression of cloned genes, in Meth. Enzymol., San Diego, CA: Academic Press. 185, 60–89.

13. Kimata,Y.,Iwaki,M.,Lim,C.R.,andKohno,K. (1997) A novel mutation which enhances the fluorescence of green fluorescent protein at high temperatures. *Biochem. Biophys. Res. Commun.* 232, 69–73.

14. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

15. Hoover, D. M., and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**.

16. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164, 49–53.

17. Lorimer, I. A. J., and Pastan, I. (1995) Random recombination of antibody single-chain Fv sequences after fragmentation with DNase I in the presence of Mn^{2+}. *Nucleic Acids Res.* 23, 3067–3068.

18. Prodromou, C., and Pearl, L. (1992) Recursive PCR: a novel technique for total gene synthesis. *Protein Eng.* 5, 827–829.

19. Bosch, B. J., Martina, B. E. E., van der Zee, R., Lepault, J., Haijema, B. J., Versluis, C., Heck, A. J. R., de Groot, R., Osterhaus, A., and Rottier, P. J. M. (2004) Severe acute respiratory syndrome coroavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. USA* 101, 8455–8460.

20. Liu, S. W., Xiao, G. F., Chen, Y. B., He, Y. X., Niu, J. K., Escalante, C. R., Xiong, H. B., Farmar, J., Debnath, A. K., Tien, P., and Jiang, S. B. (2004) 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* 363, 938–947.

21. Zhu, J. Q., Xiao, G. F., Xu, Y. H., Yuan, F., Zheng, C. Y., Liu, Y. Y., Yan, H. M., Cole, D. K., Bell, J. I., Rao, Z. H., Tien, P., and Gao, G. F. (2004) 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.* 319, 283–288.

22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.

23. Tripet, B., Howard, M. W., Jobling, M., Holmes, R. K., Holmes, K. V., and Hodges, R. S. (2004) Structural characterization of the SARS-coronavirus spike S fusion protein core. *J. Biol. Chem.* 279, 20836–20849.

24. Joo, H., Lin, Z. L., and Arnold, F. H. (1999) Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* 399, 670–673.

25. Stemmer, W. P. C. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389–391.