Indian Journal of Animal Sciences **82** (7): 665–670, July 2012

Batch isolation of novel sequences targeting regions of rapid viral variations

WENWEN YU[1], XINMING YANG[2], YE GAO[3], XIAOXIAO ZHANG[4], RAN HE[5], XIAOQIAN MA[6], MIAO LV[7], RONGFENG CUI[8], SHULI WANG[9], YONGCHANG CAO[10], SHAOPING XIE[11], WEI SHI[12], ZHIXUE WANG[13], LIANG CUI[14], YAN CAO[15], YING ZHANG[16], YUAN WANG[17], FAN FENG[18], WEIGUO CAO[19] and QIUYUN LIU[20]

Sun Yat-sen University, Guangzhou, P R China

Received: 30 July 2011; Accepted: 2 January 2012

ABSTRACT

Rapid evolution is widespread in the viral kingdom and a major concern for developing universal vaccines. The isolation of large numbers of viral sequence variants at highly variable regions in viral proteins remains a daunting challenge. Foot-and-mouth disease virus (FMDV) is a picornavirus and an underlying cause of a highly contagious disease of cloven-hoofed animals which often results in substantial economic losses. This study aimed at developing a combined method for the isolation of novel sequences to cope with rapid viral variations at the G-H loop of FMDV VP1 protein. DNA primer sets harbouring random nucleotides scattered in the G-H loop region, carrying additionally coding sequences for a T cell epitope and the carboxyl terminus, were designed and assembled by asymmetric PCR. After eliminating insert-free vector background and enriching positive clones, 100 novel sequences targeting the rapid viral variations of the G-H loop were obtained from 2 plasmid libraries. The method allowed radical changes in amino acid sequence, and our study has identified critical steps in the batch discovery of viral sequence variants, paving the way for future research on the possibility of developing a polyvalent universal vaccine.

Key words: Batch discovery, Foot-and-mouth disease virus, G-H loop, Novel sequences, Viral variations

Rapid evolution is a hallmark of many viruses and a major concern for vaccine development. Batch isolation of novel viral sequences to address viral variations is of prime importance. However, such methodologies are not well developed in terms of viral agents.

Present adderss: [1,3,5,6]Research Associate (yuwenwen2001@sina.com; gaoye.leo@gmail.com; ran.he@umassmed.edu; mxq8933@yahoo.com.cn). [2,4]Ph.D.student (ben_young@126.com; putao_grape57819@yahoo.com.cn), Cell transplantation and Gene therapy Institute, The Third Xiang Ya Hospital of Central South University, Changsha, Hunan 410013; [7]Undergraduate (bestlvmiao@sina.com); [8]Ph.D candidate (86271911@qq.com); [9]Patent Attorney (wangshuli92@126.com); [10]Professor (caoych@mail.sysu.edu.cn); [11]Research Associate (ouyangxz@singnet.com.sg); [12]Ph.D student (shiweiphd@sina.com); [13]Research Associate (wangzhixue168@163.com); [14]Research Associate (clzsu@163.com); [15]Research Associate (cybirdlllz@yahoo.com.cn); [16]Intermediary Technician (estelle_zy@yahoo.com.cn) Guangzhou Center for Disease Control and Prevention, Guangzhou 510080, P. R. China.

[17,18]Research Associate (angeleinstein@gmail.com; gunnerandff@gmail.com);[19]Professor (wgc@clemson.edu), Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, USA.

[20]Principle Scientist, Associate Professor and Incharge (liuqiuyunzhongda@yahoo.com.cn).

FMD is a destructive viral disease of cloven-hoofed animals. Animal culling, restricted animal travel and vaccination are common measures to control this disease in enzootic countries (Baxter *et al.* 2009, Dory *et al.* 2009). The highly variable G-H loop of the VP1 protein of Foot and Mouth Disease (FMDV) virus is the major antigenic site (Capozzo *et al.* 2011, Baxter *et al.* 2009). Peptide vaccines based on several antigenic sites displayed various levels of protection in guinea pigs, swine, and cattle. As an initial attempt to cope with rapid viral evolution with the ultimate goal of developing a polyvalent universal vaccine, we set out to isolate novel sequences corresponding to the G-H loop of type O FMDV VP1 proteins *via* random DNA technology.

MATERIALS AND METHODS

Kits and reagents: DNA fragments were recovered from agarose gel slices using gel extraction kits according to manufacturer's instructional manual.

Media and culture condition: LB/glucose/glycerol medium (LBGG) was supplemented with 2% w/v glucose and 4% v/v glycerol in addition to LB (Sambrook *et al.* 1989). Ampicillin was added at a concentration of 100 g/ml. Antioxidant oligomeric proanthocyanidins was dissolved in absolute ethanol and used at a final concentration of 10 mg/l.

Table 1. The chimeric peptides and encoding DNA sequences of the random and semi-random designs[a]

Semi-random design	ETQTQRRHHTDVAFVLDRFRHKQPLIAPAKQLLPPS KY(G/S/A)(D/T/V/G/E/Q)(T/A/G/S)(S/R/H/P/A/L)(T/V/A/L)(N/S/T/A)(N/K/T)VRGDL(Q/R)VL(A/T)QK(A/T)(E/A/T)R(T/A/P/S)LP
random design [b]	ETQTQRRHHTDVAFVLDRFRHKQPLIAPAKQLLPPS KYXXXXXXXVRGDL(Q/R)VL(A/T)QK(A/T)(E/A/T)R(T/A/P/S)LP
Designed semi-random coding sequence	GAAACTCAAACTCAAAGAAGACATCATACTGATGTCGCTTTCGTTTTGGATAGATTTGTAGAC ATAAACAGCCATTGATTGCTCCAGCTAAACAATTGTTGCCACCATCTAAGTACKSTSWW DSTVNWDYKNMTAMWGTCAGAGGTGATTTGCRAGTTTTGRCTCAAAAARCTGMW AGANCWTTGCCATGA
Designed random coding sequence	GAAACTCAAACTCAAAGAAGACATCATACTGATGTCGCTTTCGTTTTGGATAGATTTGT TAGACATAAACAGCCATTGATTGCTCCAGCTAAACAATTGTTGCCACCATCTAAGTACVNNVN NVNNVNNVNNVNNVNNGTCAGAGGTGATTTGCRAGTTTTGRCTCAAAAARCTGMWA GANCWTTGCCATGA

[a]M=A/C, R=A/G, W=A/T, S=G/C, Y=C/T, K=G/T, V=A/G/C, H=A/C/T, D=A/G/T, B=G/C/T, N=A/G/C/T; [b]X: any amino acid.

Viral variation targeted sequence isolation via random DNA technology: Sequence alignment of the G-H loops of the VP1 proteins from type O strains was made from the proteins retrieved from GenBank database (Online supplement). The consensus sequence is embedded in the semi-random design in Table 1. The peptide E and F was from FMDV C3 Arg85 (Taboga *et al.* 1997). Tripeptide PPS served as a hinge between peptide F and peptide G. The functional diversity of the peptide libraries of the random and semi-random designs is 2.57×10^{10} and 3.32×10^{7}, respectively. The 2 libraries are denoted as random library and semi-random library hereafter (Table 1; Fig.1).

Five primers were assembled in a step-wise fashion. P2/P3 and P4/P5 (supplementary Table 1) were assembled via asymmetric PCR (Gyllensten and Erlich 1988) with a weight ratio of 15:1 and 1:15 respectively, using Pfu DNA polymerase in the presence of 2.5% 2-Pyrrolidone (Chakrabarti and Schutt 2001). PCR mixture was heated at 94°C for 2 min, followed by 26 cycles of incubation at 94°C for 20sec, 35°C for 45sec and 72°C for 35sec, and a final extension at 70°C for 5 min, which gave rise to P23 and P45 subassembly products.

PCR reaction mixtures were mixed in a weight ratio of P23: P45=1:4 and amplified using the aforementioned conditions except with a 30°C annealing temperature, giving rise to a 174 bp secondary assembly product P25. The P1 and P25 reactions were combined in approximately equal molar ratio and amplified to generate a 220 bp tertiary assembly product P15 with the following conditions: PCR mixture was heated at 94°C for 2 min, followed by 30 cycles of incubation at 95°C for 30sec, 27°C for 45sec and 72°C for 30sec, and a final extension at 70°C for 5 min.

Layout of the chimeric peptides: E—F— PPS—G
sequence of E peptide (T-cell epitope): ETQTQRRHHTDVAFVLDRF
Sequence of F peptide (C-teminus): RHKQPLIAPAKQLL
G peptide (B-cell site): homologues of G-H loop of Vp1 capsid protein

Fig 1. The layout and amino acid sequences of the chimeric peptides.

For the random design, we replaced P4 with P4b and repeated the assembly process to generate a 220 bp tertiary assembly product P15b. The 457 bp chimeric fragments encoding *S. cerevisiae* a-factor secretion signal and combinatorial peptides were amplified from 40 ng of pPICZaA and 50 ng of P15 or P15b to yield final assembly products Pa5 and Pa5b respectively, using P6 and P5 primers with the following conditions: PCR mixture was heated at 94°C for 2 min, followed by 30 cycles of incubation at 94°C for 20sec, 37°C for 45sec and 72°C for 35sec, and a final extension at 70°C for 5 min.

The Pa5 and Pa5b amplicons were gel purified respectively and double digested with 40 U each of EcoRI and XbaI for 4 h. After heat inactivation at 70°C for 15 min, the fragments were precipitated with ethanol and dissolved in 30ml distilled water. The *S. cerevisiae* shuttle vector pYES2/CT was triple digested with 40 U each of EcoRI, XhoI and XbaI overnight along with 40 U of CIAP (calf intestinal alkaline phosphotase). The enzymes were subsequently heat inactivated and the mixture was precipitated with ethanol.

350 ng of Pa5 fragments or 350 ng of Pa5b fragments were ligated with pYES2/CT respectively at an approximate molar ratio of 10:1 in the refrigerator (5–8°C). The ligation mixtures were precipitated with ethanol and dissolved in 10 ml sterile double distilled H_2O or less, and electroporated to *Escherichia coli* ElectroMax-DH10B strain (Wang *et al.* 2007). Cells were plated to LBGG/ampicillin supplemented with oligomeric proanthocyanidins at a final concentration of 10 mg/L. Thousands of colonies appeared the next day. Single colonies were inoculated to LBGGO (LBGG/oligomeric proanthocyanidins/ampicillin) culture, and then limited numbers of plasmid isolations were unsuccessful as most clones may have very low amount of plasmid DNA.

The clones on the plates were enriched as described below. Bacteria on the plates were washed off with sterile water and propagated in liquid LBGGO culture at 30°C with shaking. The plasmid DNA was then isolated and treated

with 29 U NotI and 8 U CIAP overnight to linearize the circular insert-free vector DNA. After electroporation, transformants were spread on LBGGO plates at 37°C, and inoculated into LBGGO liquid media and cultured at 30°C with shaking. Some clones were subjected to PCR characterization with P7 and P8, or P7 and P3, or P7 and P5, or P6 and P5, giving rise to 730 bp, 530 bp, 580 bp and 440 bp fragments respectively. Eighty % of the clones showed fragments with expected sizes. Randomly picked clones were sequenced with T7 primer, and positive clones were re-sequenced in the opposite orientation with P9 primer. Initial sequencing was performed on about 200 clones on a ABI 3730XL sequencer.

All sequences had a Qv values of at least 30, with exceptions on a few clones which had several bases with a Qv values around 10 to 30 that were attributed to homo-nucleotide tracts. Qv values of 30, 20 and 10 correspond to an error probability of 0.1, 1 and 10%, respectively. Most positive clones could be sequenced using DNA extracted from cells propagated on media free of oligomeric proanthocyanidins.

The colonies from the random library were washed off and pooled, followed by PCR amplification and sequencing with P9 primer. The sequencing chromatogram of the PCR amplicons from pooled clones of the random library was assessed for library complexity.

Second phase sequencing and DNA library evaluation: To evaluate yield of positive clones of the 2 libraries, the combinatorial library stocks (20 ml) were plated on LBGGO media, and single colonies were inoculated to 2X YT media free of oligomeric proanthocyanidins for growth at 37°C. Second phase sequencing was conducted with T7 primer and P9 primer on ABI 3730XL sequencer by BGI Shenzhen, China. Clonal representation may have changed after 2-year storage in the freezer as more toxic clones tended to die more quickly. Therefore, only the second-phase sequencing was considered for mathematical calculations.

DNA clones were named as Y series and ZX, and the corresponding encoded peptides were denoted as YWW series and ZXX. Y32 to Y94 were from the semi-random library, and Y95 to Y100 were from the random library. The Qv values were generally over 30 on second-phase sequencing in FMDV VP1 fusion gene regions.

Statistical analysis: SPADE calculations using different models were made using software downloaded from (http://chao.stat.nthu.edu.tw/softwareCE.html).

Sequence alignment: Raw sequences were translated into amino acid sequences using DNAClub. Sequence alignments with default parameters were performed by Clustal X version 2.0.10 and displayed by bioedit 7.0.9.0.

Sequence alignment: Sequence alignment of the G-H loops of the VP1 proteins from type O strains was made from the following proteins retrieved from GenBank database (http://www.ncbi.nlm.nih.gov/): AJ294914, AJ303520, AJ294918, AJ303521, AJ296328, AJ004663, AJ303522, AJ303484, AJ303523, AJ303483, AJ318833, AJ318844, AJ318847, AJ303525, AJ303526, AJ318848, AJ131469, AF095884, AF095885, AF026168, AJ296322, AJ294923, AJ318836, AJ294925, AJ294922, AJ294924, AF525458, AJ131468, AJ131470, AJ294917, AJ294913, AJ294915, AJ294916, AJ294919, AJ294920.

Sequence accession numbers: DNA sequences and encoded peptide sequences have been deposited in Genbank and are accessible under the codes JN088739 to JN088838.

RESULTS AND DISCUSSION

Principle of the random DNA technology

Two 5-primer sets were designed after the initial sequence alignment (Table 1 and supplementary Table 1). DNA fragments were assembled by asymmetric PCR with high fidelity DNA Polymerase Pfu, and attached downstream to

Supplementary Table 1. Primers

P1 For DNA fragment assembly	TCTCTCGAGAAAAGAGAAACTCAAACTCAAAGAAGACATCA TACTGATGTCGCTTTCGT
P2 For DNA fragment assembly	CATACTGATGTCGCTTTCGTTTTGGATAGATTTGTTAGACATAAACA GCCATTGATTGC
P3 For DNA fragment assembly	GTACTTAGATGGTGGCAACAATTGTTTAGCTGGAGCAATCAA TGGCTGTTTATGTC
P4 For DNA fragment Pα5 assembly	TGTTGCCACCATCTAAGTACKSTSWWDSTVNWDYKNMTAMWG TCAGAGGTGATTTG
P5 For DNA fragment assembly	CCTCTAGATCATGGCAAWGNTCTWKCAGYTTTTTGAGYCAAAACTYG CAAATCACCTCTTGAC
P4b For DNA fragment Pa5b assembly	TGTTGCCACCATCTAAGTACVNNVNNVNNVNNVNNVNNVNNG TCAGAGGTGATTTG
P6α-factor secretion signal 5' primer	GCGAATTCAAAAATGAGATTTCCTTCAA
P7 Upstream primer	CCTCTATACTTTAACGTCAAGGAGAAAAAACCCCGGATCG
P8 Downstream primer	GGCGTGAATGTAAGCGTGACATAACTAATTACATGATGCG
P9 Downstream sequencing primer 1	TAGGATCAGCGGGTTTAAACTC
P10 Downstream sequencing primer 2	GATGGTGATGATGACCGGTACG
T7 Upstream sequencing primer	TAATACGACTCACTATAGGG

the Kex2 cleavage site of the *Saccharomyces cerevisiae* α-factor secretion signal fragment, and inserted into the EcoRI and XbaI restriction sites of yeast shuttle vector pYES2/CT. A simultaneous treatment of the vector with EcoRI, XbaI, and XhoI removed most insert-free vector DNA by a complete digestion of the vector DNA to a poorly transformed linear form (Ouyang *et al.* 2006). Recircularization of the insert-free linear molecules was prevented by CIAP treatment a priori (Ouyang *et al.* 2006). The clones carrying FMDV sequences were obtained by using an ultra-efficient electroporation method. Electroporated cells were plated on LBGG medium supplemented with potent antioxidant oligomeric proanthocyanidins to reduce the toxicities of the expressed peptides. After pooling the clones, plasmid DNA was isolated and treated with NotI and CIAP to further remove clones with insert-free vectors, and re-electroporated to eliminate clones bearing few plasmids. Eventually full length VP1 gene could be constructed.

The nucleotide sequence of the constant region of the chimeric peptides was designed according to the codon bias of *S. cerevisiae*. The random design adopted VNN triplets to avoid stop codons and also eliminated aromatic residues and cysteine, since the natural G-H loop sequences tended to be devoid of aromatic amino acid residues.

Evaluation of the random library

All the completely random sites by design were fully degenerated in the chromatograms of the pooled clones (Fig. 2), thus showing low Qv values. There were 15 fully degenerate N sites in the random design. The probability was 3/4 to deselect G once and we needed multiply 3/4 once for each deselection with an additional clone. To allow for full coverage by the 4 bases in the 15 N sites, we needed to have a probability less than 1/30 (<0.5×1/15) for reasonable assurance. For approximation, we multiplied 4 by $(3/4)^{17}$ which was a bit smaller than 1/30, so we had at least 17 potential positive clones in the random library that made the random sites in the chromatogram fully degenerate. Probabilities for simultaneously deselecting 2 or 3 bases, respectively, were too small and neglected. The rest of the random sites were less degenerate in design and fewer in number, and therefore they were fully covered by the aforementioned probability.

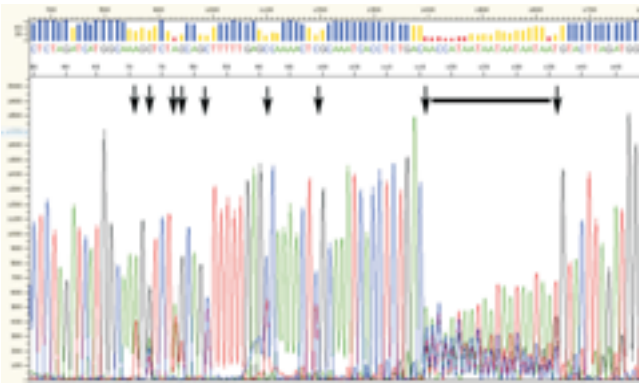

Fig. 2. The sequencing chromatogram of the PCR amplicons amplified from the pooled clones of the random library. Fully random sites are indicated with arrows. The stretch of random sites included 7 VNN codons and PCR amplicons were sequenced in a reverse orientation.

Estimation of the output from additional sequencing of the semi-random library

To establish formulas for evaluating output of such libraries, the second-phase sequencing was performed on the 2-year-old library stocks. To estimate the yield from additional sequencing in the remaining library in consideration for both productivity and cost effectiveness, phase 2 was divided into phase 2a and 2b, and the following formula was proposed. Assuming there are D, duplicated clones in the initial sequenced pool of T1 size, additionally T2 clones are to be sequenced, and R represents the average redundancy rate in the (T1+T2) pool, then we have the following approximation:

When $R \leq 2$

$$D=T_1\times(T_1/R)/[(T_1+T_2)/R]\times(R-1)/R=T_1^2\times(R-1)/[(T_1+T_2)R] \quad [1]$$

T1 × (R-1)/R in [1] represents the duplicated sub-pool in T1 pool based on the redundancy rate in (T1+T2). However, this duplicated sub-pool corresponds to distinct sequences in the non-redundant [(T1+T2)/R] pool, and the genuine number of duplicated clones in T1 is T1/(T1+T2)×T1 × (R-1)/R. Therefore, we can calculate how many novel clones we can acquire from next round of sequencing. Substantially overrepresented clones may be removed from the dataset for achieving higher accuracy. The derivations and mathematical formulae for higher redundancies can be found in the Online Supplement.

In the second-phase sequencing for the semi-random library, we initially obtained 25 positives with 1 duplicate (supplementary Table 2), which would be rounded from probable events anywhere from 0.50 to 1.49 duplicates. The duplicates were projected to be between 6 and 18 in 86 total positives obtained in phase 2a and 2b (Table 2), which was in accordance with the final number of 16 duplicates. Approximately 35 or 50 new positive sequences were expected to be captured from the 86 or 172 positive clones from future sequencing, respectively. Our estimates were largely consistent with SPADE predictions. However, the statistical software based on Multinomial and Poisson models, failed to provide the initial calculations when there was only a small sample size and single duplicate. Our formulae suggest that a cost effective and productive sequencing is mostly done at a redundancy rate of R<2. The mathematical formulae of non-linear regression for estimating future sequencing output are close approximations and easy to use. The summary of the first- and second-stage sequencing can be found in the Online Supplement.

Table 2. The evaluation of second phase sequencing of isolated clones of the semi-random library [a].

	Successful sequenced/ attempted	Positive rate	Number of duplicates	Number of duplicates (mutations outside of the FMD region excluded [c])	Projected duplicates from phase 2a
Phase 2a	74/75	25/74 = 33.8%	1[b]	1	
Phase 2b	170/197	61/170 = 35.3%	15[b] (among 2b[e] and between 2b[e] and 2a[e])	16 [c] (among 2b[e] and between 2b[e] and 2a[e])	6-18 (among 2b[e] and between 2b[e] and 2a[e]) [d]

[a] All the second-phase sequencing was conducted on DNA isolated from cells propagated in media free of oligomeric proanthocyanidins at 37°C; [b] The redundant clones include 14 doublets and 1 triplets; [c] The sequence difference in the *Saccharomyces cerevisiae* mating type α secretion signal region between a pair of clones are neglected thus yielding a doublets; [d] See 3.3; [e] Phase 2a or 2b

Next generation or third generation sequencing can be integrated into our method. However, the post-discovery sequence re-assembly *via* gene synthesis and re-sequencing could be time consuming, labour intensive and costly. Yet, the high throughput method can provide us information on the overrepresentation of some clones, which are likely to be expressed efficiently in *E. coli* or other expression systems if recombinant protein vaccine is to be produced.

A few critical points need to be considered for a successful run of this procedure. First, asymmetric PCR should be adopted for DNA fragment assembly. It markedly reduces annealing of complementary strands, thus increasing the chances of association of overlapping fragments, which is particularly useful for assembly of random DNA fragments. The linearization of plasmid pPICZαA with restriction enzyme EcoRI was beneficial for the construction of chimeric fragments encoding *S. cerevisiae* α-factor secretion signal and combinatorial peptides, for it reduced rapid reassociation of circular pPICZαA molecules after denaturation. Second, since clones bearing insert-free vectors were prevalent in the initial libraries as a result of the infinitesimal incomplete digestion and subsequent super-efficient transformation, a simple NotI/CIAP treatment during restriction digestion, or after ligation, or post-transformation plasmid isolation was capable of removing these vectors a priori. It worked by the complete linearization via a NotI digestion and prevention of recircularization of the insert-free linear vector molecules via CIAP (Ouyang *et al*. 2006). NotI (or previously mentioned XhoI) site is located between EcoRI and XbaI in the pYES2/CT vector, and its digestion does not interfere ligation. NotI, when combined with CIAP, turned out to be markedly more efficient in eliminating insert-free vector harbouring clones than other restriction enzymes, which may be attributed to its all G/C recognition sequence that has stabilized the secondary structure of the DNA and rendered it completely accessible to the restriction enzyme. This treatment has drastically increased the positive rate (data not shown), thus enhancing cost effectiveness during the sequencing stage. This method eliminated the complicated screening steps such as transfection, plaque purification, affinity selection and other tedious process in peptide and protein display technologies for obtaining sequence variants. The selection of positives *via* inactivation of a-complementation did not work because sustained high-level expression selected against many random positive clones (data not shown). Third, an ultra-efficient electroporation protocol and a super-efficient *E. coli* strain were prerequisites for this method as the vast majority of transformants harbouring random sequences were toxic to *E. coli* host and would not form colonies. *E. coli* strain ElectroMAX™ DH10B™ turned out to suit this purpose, and it showed over 10-fold higher electroporation efficiencies than XL1-Blue MRF' cells. The supplement of oligomeric proanthocyanidins may be beneficial for reducing DNA rearrangement background and representation in the libraries thus increasing the output rate of positive clones.

Current vaccines for FMDV are inactivated forms of the whole virions and are serotype-specific, which creates opportunities for viral escape from containment facilities. Success with peptide vaccine on cattle has been limited (Taboga *et al.* 1997) for two reasons. Firstly, the inability of the peptides to present one or multiple native conformations (Capozzo *et al*. 2011); Secondly, the diversity of polymorphisms on the major histocompatibility complex regions and other genetic components present on the cattle (Leach *et al.* 2010, Baxter *et al.* 2009). Sera from calves vaccinated with expressed proteins or DNA constructs of vesicular stomatitis virus glycoprotein G carrying a tandem dimer of type C FMDV antigenic site A (ASA), efficiently immunoprecipitated whole virions and neutralized viruses in cell culture (Capozzo *et al.* 2011). This result suggests that the native conformations of the antigenic site have been displayed in the fusion proteins. Modeling of the protein structure has indicated the presence of an exposed configuration of one ASA peptide displayed on the surface of the molecule. The report corroborated previous findings that a dimer or a tetramer of the antigenic site or as fusions to carrier proteins was more immunogenic than the monomer (Dory *et al.* 2009, Fan *et al.* 2007, Brown 1988). Some bovine MHC *DRB3* Alleles and their binding pockets were strongly associated with IgG1 and IgG2 immune responses to a 40-mer peptide consisting of the G-H loop region and the C-

terminal residues of the VP1 protein (Baxter *et al.* 2009). This work also explains in part the lack of solid protection by the monomeric antigenic site in cattle, and facilitates the design of an effective peptide or DNA vaccine against FMDV based on MHC polymorphisms. A study on quantitative trait loci for variation in immune response to a FMDV peptide has found that the majority of QTL located outside the MHC locus (Leach *et al.* 2010). Another study has discovered that Hsp70 improved the presentation of a 25-mer FMDV peptide to CD4+ T cells via the assay of T cell proliferation (McLaughlin *et al.* 2010). One future direction will be the design of a DNA vaccine with our novel G-H loop sequences in a multimeric fashion fused with a carrier gene, and test its efficacy on animals. Many amino acid residues in the most variable region of our G-H loop sequences are hydrophilic, which are likely to be displayed on the surface of the chimeric molecules as antigens.

This paper has described a rapid, cost-effective and robust method to generate novel sequences. It may be the first feasible approach to the batch isolation of radical viral sequence variants that paves the way for future investigations.

ACKNOWLEDGEMENTS

This work was supported by Guangdong Science and Technology Program (No. 2008B020100001), Guangdong Natural Science Foundation of China (S2011010004264), National Natural Science Foundation of China (No. 30370799), Wen's Group (wens-kjhz-2007-03-09-01) and Opening Fund of Laboratory at Sun Yat-sen University (KF201028) and (KF201125). Moreover, we thank Yan Shi for language help and anonymous reviewers for valuable suggestions.

REFERENCES

Baxter R, Craigmile S C, Haley C, Douglas A J, Williams J L and Glass E J. 2009. BoLA-DR peptide binding pockets are fundamental for foot-and-mouth disease virus vaccine design in cattle. *Vaccine* **28**(1): 28–37.

Brown F. 1988. Use of peptides for immunization against foot-and-mouth disease. *Vaccine* **6**:180–82.

Capozzo A V, Wilda M, Bucafusco D, de los Ángeles Lavoria M, Franco-Mahecha O L, Mansilla F C, Pérez-Filgueira D M and Grigera P R. 2011. Vesicular Stomatitis Virus glycoprotein G carrying a tandem dimer of Foot and Mouth Disease Virus antigenic site A can be used as DNA and peptide vaccine for cattle. *Antiviral Research* **92**(2): 219–27.

Chakrabarti R and Schutt C E. 2001. The enhancement of PCR amplification by low molecular weight amides. *Nucleic Acids Research* **29**(11): 2377–81.

Dory D, Rémond M, Béven V, Cariolet R, Backovic M, Zientara S and Jestin A. 2009. Pseudorabies virus glycoprotein B can be used to carry foot and mouth disease antigens in DNA vaccination of pigs. *Antiviral Research* **81**: 217–25.

Fan H, Tong T, Chen H and Guo A. 2007. Immunization of DNA vaccine encoding C3d-VP1 fusion enhanced protective immune response against foot-and-mouth disease virus. *Virus Genes* **35**: 347–57.

Gyllensten U B and Erlich H A. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of National Academy of Sciences USA* **85** (20): 7652–56.

Leach R J, Craigmile S C, Knott S A, Williams J L and Glass E J. 2010. Quantitative trait loci for variation in immune response to a Foot-and-Mouth Disease virus peptide. *BMC Genetics* **11**:107–21.

McLaughlin K, Seago J, Robinson L, Kelly C and Charleston B. 2010. Hsp70 enhances presentation of FMDV antigen to bovine CD4+ T cells in vitro. *Veterinary Research* **41**(3): 36–48.

Ouyang X Z, Yu W, Wang Q, He J and Liu Q. 2006. A novel assay to quantitate in vivo perfect recircularization rate of restriction enzyme-generated ends. *Journal of Rapid Methods and Automation in Microbiology* **14**: 283–90.

Sambrook J, Fritsch E F and Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Taboga O, Tami C, Carrillo E, Núñez J I, Rodríguez A, Saíz J C, Blanco E, Valero M L, Roig X, Camarero J A, Andreu D, Mateu M G, Giralt E, Domingo E, Sobrino F and Palma E L. 1997. A large-scale evaluation of peptide vaccines against foot-and-mouth disease: lack of solid protection in cattle and isolation of escape mutants. *Journal of Virology* **71**(4): 2606–14.

Wang S L, Luo Y, Yi X, Yu W, Xu Z, Ma X, He J and Liu Q. 2007. A highly efficient and highly reliable protocol for transformation of *E. coli* by electroporation. *Journal of Rapid Methods and Automation in Microbiology* **15**: 253–58.