Structure of transmembrane subunits gp47 of the foamy virus envelope glycoproteins

M. WANG[1], H. ZHANG[1], Q.-M. LIU[1], Y. SUN[1*], Z. LI[1*], W.-H. LIU[2], X.-H. HE[2], J. SONG[1], Y.-X. WANG[1]

[1]College of Life Sciences, Shaanxi Normal University, Xi'an 710062, P. R. China; [2]Pathogenic Organism and Infectious Diseases Research Institute, School of Basic Medical Sciences, Wuhan University, Wuhan 430071, P. R. China

Summary. – The successful foamy viruses (FVs) infection includes at least two essential events, attachment to the cell surface and fusion of the viral envelope with the cell membrane. For the FVs, membrane fusion between virus and cell is mediated by envelope glycoprotein (Env) transmembrane (TM) subunit gp47. Compared with other retroviruses, FV TM subunit shares a similar but not identical structural characteristic. This paper focuses on *in sillico* analyses of all 15 available FV TM subunits gp47 based on their amino acid sequences. The hydrophobicity analysis revealed that the 15 FVs gp47 had two prominent hydrophobic regions, the N-terminal fusion peptide (FP) and the C-terminal region, which included a membrane-spanning domain (MSD) and a membrane proximal ectodomain region (MPER). In most FVs gp47, two heptad repeats, the coiled coils characterized by repetition of 7-amino acid-motif, were found to be correspondingly located downstream of FP (named "N-HR") and the upstream of MPER (named "C-HR"). Furthermore, the solvent accessibility and secondary structure were predicted for all FVs gp47. These observations suggested that FVs gp47 possessed several fusion domains, which were necessary in the process of lipid membrane fusion between FVs and the target cells.

Keywords: foamy virus; envelope glycoprotein; transmembrane subunit gp47; hydrophobicity analysis; fusion domain

Introduction

Foamy viruses (FVs), also called spumaretroviruses, are complex retroviruses, which constitute the only genus in

*Corresponding authors. E-mail: lizhi@snnu.edu.cn; phone: +86-29-85310581.

Abbreviations: BFV = bovine foamy virus; BSV = bovine syncytial virus; C-HR = C-terminal heptad repeat; EFV = equine foamy virus; Env = envelope glycoprotein; ERRS = endoplasmic reticulum retrieval signal; FFV = feline foamy virus; FP = fusion peptide; FSV = feline syncytial virus; FV(s) = foamy virus(es); HBV = hepatitis B virus; HIV = human immunodeficiency virus; HR = heptad repeat; MMTV = mouse mammary tumor virus; MPER = Membrane proximal ectodomain region; MSD = membrane-spanning domain; N-HR = N-terminal heptad repeat; PFV = prototype foamy virus; SFV$_{agm}$ = african green monkey simian foamy virus; SFV$_{cpz}$ = chimpanzee foamy virus; SFV$_{gor}$ = gorilla foamy virus; SFV$_{mac}$ = macaque foamy virus; SFV$_{mar}$ = marmoset foamy virus; SFV$_{spm}$ = spider monkey foamy virus; SFV$_{sqm}$ = squirrel monkey foamy virus; SIV = simian immunodeficiency virus; SloEFV = sloth endogenous foamy virus; SU = central surface domain; TM = transmembrane domain

the subfamily *Spumaretrovirinae* of the family *Retroviridae* (Dirk and Axel, 2011). Till now, 15 FVs have been reported, which were derived from 6 non-primate and 9 primate hosts, including cat, horse, cattle, non-human primates and human. As indicated by their names, FVs possess highly fusogenic activity, which induces infected cells *in vitro* to form foam-like multinucleated cells. FVs can establish lifelong persistent infections in their hosts in the absence of any pathogenicity (Linial, 2000). As with other retroviruses, the FV envelope protein (Env) harbors essential features that allow viral adsorption, uptake and fusogenic release of capsids into the cytoplasm. However, FV Env protein has a number of features that are not typical. The most prominent feature is that the expression of the spumaretroviral Env protein in infected cells induces massive syncytium formation *in vitro*.

FV Env full-length precursor protein, which belongs to type III transmembrane protein, is targeted to the rough endoplasmic reticulum (ER) by an N-terminal leader peptide (LP). The Env precursor is then cleaved into LP (gp18), central surface (SU, gp80) and transmembrane subunit (TM, gp47) by furin or furin-like cellular proteases during the

transport to the cell surface (Duda *et al.*, 2004; Lindemann and Rethwilm, 2011; Sun *et al*, 2012). Proteolytic cleavage of SU and TM subunits is essential for viral infectivity (Duda *et al.*, 2004; Lindemann and Rethwilm, 2011). The SU subunit mediates the binding of viruses to the cellular receptors, whereas the TM subunit gp47 is involved in the fusion of the lipid membranes between FVs and the target cells (Pietschmann *et al.*, 2000; Peisajovich and Shai, 2003). Although some features of several FVs gp47 have been analyzed (Wang and Mulligan, 1999; Sun *et al.*, 2012), the fusogenic TM subunits gp47 of all 15 available FV Envs are not fully dissected. In order to better understand the features of FVs gp47, we used computer-based methods to perform sequence analyses and predictions focused on all 15 available FV gp47 sequences. The results obtained in this work would stimulate and guide future molecular studies of FV glycoproteins and better understanding the virus Env-mediated cell fusion.

Materials and Methods

The sequences involved in this article were obtained directly by searching the NCBI Nucleotide database and Protein database (Access Nos. are summarized in Table 1). The gp47 sequence of Sloth Endogenous Foamy Virus (SloEFV) was deduced from a previous paper (Aris *et al.*, 2009). The predictions of solvent accessibility and secondary structure of gp47 were made with the online program Predictprotein (https://www.predictprotein.org/home). The hydrophobicity analysis was made by the ExPASy Proteomics Server using the scale Hphob/Kyte & Doolittle (http://web.expasy.org/cgi-bin/protscale/protscale.pl)(Gasteiger *et al.*, 1999) and the Membrane Protein Explorer (MPEx) program from the Stephen White laboratory at UC Irvine (http://blanco.biomol.uci.edu/mpex/)(Snider *et al.*, 2009). Helical wheel analysis was made by the HeliQuest program (http://heliquest.ipmc.cnrs.fr/) (Gautier *et al.*, 2008). The membrane-spanning domain (MSD) was predicted by the online program HMMTOP from Hungarian Academy of Sciences (http://www.enzim.hu/hmmtop/html/adv_submit.html) (Tusnady and Simon, 1998). The prediction of the heptad repeats was made by the COILS program (http://www.ch.embnet.org/software/COILS_form.html) (Lupas *et al.*, 1991) and the LEARNCOIL-VMF program (http://groups.csail.mit.edu/cb/learncoil-vmf/cgi-bin/vmf.cgi) (Berger and Singh, 1997). The alignments of multiple sequences were analyzed using the Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Goujon *et al.*, 2010).

Results

Solvent accessibility and secondary structure of FVs gp47

According to the possible solvent accessibility assays, more than 50% residues of the FVs gp47 were buried (Table 1). Most of the buried residues were scattered, except that there was a buried segment in the C-terminus, which was more likely to be a membrane-spanning domain (MSD) (supplementary Fig. 1). The secondary structures of FVs gp47 were predicted as well. The data revealed that there were three structural regions within gp47 subunits: two long

Table 1. Solvent accessibility and secondary structure composition of FVs gp47

Virus	Acc. No. of genome sequence	Host species	Solvent accessibility			Secondary structure composition		
			Buried (%)	Intermediate (%)	Exposed (%)	Helix (%)	Strand (%)	Loop (%)
PFV	U21247	Human	54.68	8.87	36.45	45.08	15.11	39.81
SFVagm	NC_010820	African green monkey	54.70	8.92	36.39	43.13	16.63	40.24
SFVcpz	U04327	Chimpanzees	54.68	8.15	37.17	44.36	14.87	40.77
SFVgor	HM245790	Gorilla	54.20	9.59	36.21	44.60	15.35	40.05
SFVmac	NC_010819	Macaque	54.57	9.13	36.30	45.67	15.87	38.46
SFVmar	GU356395	Marmoset	54.37	9.47	36.17	43.45	16.75	39.81
SFVorg	AJ544579	Orangutan	54.44	8.87	36.69	45.08	14.63	40.29
SFVspm	EU010385	Spider monkey	53.73	8.92	37.35	44.34	16.63	39.04
SFVsqm	GU356394	Squirrel monkey	53.88	8.74	37.38	43.69	16.50	39.81
BFV	NC_001831	Bovine	53.11	8.85	38.04	45.93	15.31	38.76
BSV	U94514	Bovine	53.11	8.85	38.04	45.93	15.31	38.76
EFV	AF201902	Equine	52.96	9.22	37.83	45.15	15.37	39.48
FFV	NC_001871	Feline	53.94	9.07	36.99	45.35	16.47	38.19
FSV	U85043	Feline	53.46	9.31	37.23	44.39	15.99	39.62
SloEFV	\	Sloth	53.98	9.88	36.14	44.10	16.63	39.28

(a)

(c)

(b)

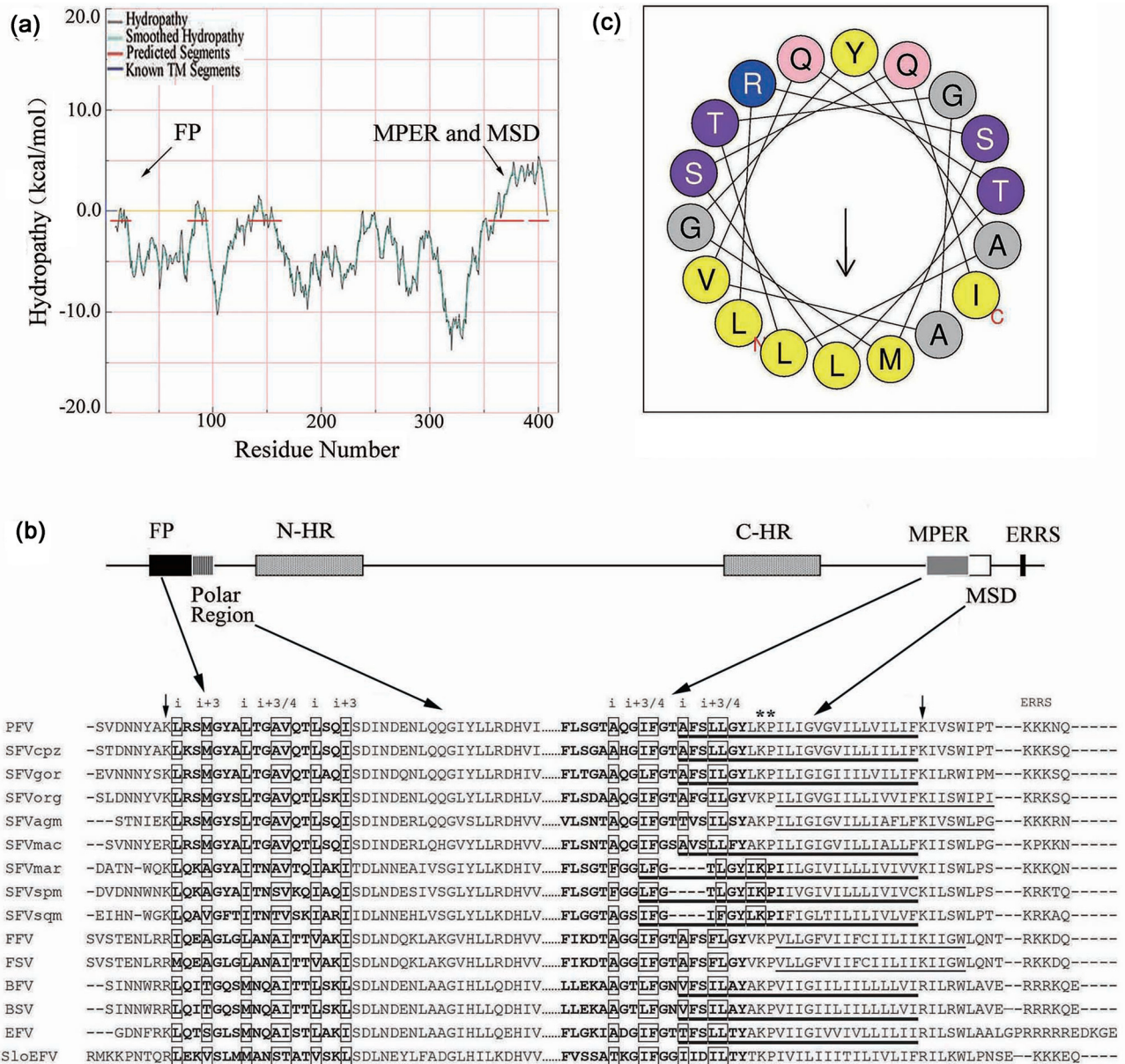|	↓ i i+3 i i+3/4 i i+3		i i+3/4 i i+3/4		ERRS
PFV	-SVDNNYAK**L**RS**M**GY**A**L**T**GA**V**QT**L**SQ**I**SDINDENLQQGIYLLRDHVI	**FLSGTA**QG**IFGTA**FS**LL**GY**LKPILIGVGVILLVILIF**KIVSWIPT	---KKKNQ-----
SFVcpz	-STDNNYAK**L**KS**M**GY**A**L**T**GA**V**QT**L**SQ**I**SDINDENLQQGIYLLRDHVI	**FLSGAA**HG**IFGTA**FS**LL**GY**LKPILIGVGVILLIILIF**KIVSWIPT	---KKKSQ-----
SFVgor	-EVNNNYSK**L**RS**M**GY**A**L**T**GA**V**QT**I**AQ**I**SDINDQNLQQGIYLLRDHIV	**FLTGAA**QG**LFGTA**FS**IL**GY**LKPILIGIGIIILVILIF**KILRWIPM	---KKKSQ-----
SFVorg	-SLDNNYVK**L**RS**M**GY**S**L**T**GA**V**QT**L**SK**I**SDINDENLQQGLYLLRDHLV	**FLSDAA**QG**IFGTA**FG**IL**GY**VKPILIGVGIILLIVVIF**KIISWIPI	---KRKSQ-----
SFVagm	---STNIEK**L**RS**M**GY**S**L**T**GA**V**QT**L**SQ**I**SDINDERLQQGVSLLRDHVV	**VLSNTA**QG**IFGTT**VS**IL**SY**AKPILIGIGVILLIAFLF**KIVSWLPG	---KKKRN-----
SFVmac	--SVNNYER**L**RS**M**GY**A**L**T**GA**V**QT**L**SQ**I**SDINDERLQHGVYLLRDHVV	**FLSNTA**QG**IFGSA**VS**LL**FY**AKPILIGIGVILLIALLF**KIISWLPG	---KPKKN-----
SFVmar	-DATN-WQK**L**QK**A**GY**A**I**T**NA**V**TQ**I**AK**I**TDLNNEAIVSGIYLLKDHIV	**FLSGTF**GG**LFG**----**TL**GY**IKPI**ILGIVILLLIVIVVK**IISWLPS	---KKKQN-----
SFVspm	-DVDNNWNK**L**QK**A**GY**A**I**T**NS**V**KQ**I**AQ**I**SDLNDESIVSGLYLLRDHVV	**FLSGTF**GG**LFG**----**TL**GY**IKPI**IVGIVILLLIVIVCK**ILSWLPS	---KRKTQ-----
SFVsqm	-EIHN-WGK**L**QA**V**GF**T**I**T**NT**V**SK**I**AR**I**IDLNNEHLVSGLYLLKDHLV	**FLGGTA**GS**IFG**----**IF**GY**LKPI**FIGLTILILIVLVFK**ILSWLPT	---KRKAQ-----
FFV	SVSTENLRR**I**QE**A**GL**G**L**A**NA**I**TT**V**AK**I**SDLNDQKLAKGVHLLRDHVV	**FIKDTA**GG**IFGTA**FS**FL**GY**VKPVLLGFVIIFCIILIIKIIGWLQNT	--RKKDQ-----
FSV	SVSTENLRR**M**QE**A**GL**G**L**A**NA**I**TT**V**AK**I**SDLNDQKLAKGVHLLRDHVV	**FIKDTA**GG**IFGTA**FS**FL**GY**VKPVLLGFVIIFCIIIIKIIGWLQNT	--RKKDQ-----
BFV	--SINNWRR**L**QI**T**GQ**S**M**N**QA**I**TT**L**SK**L**SDLNDENLAAGIHLLQDHIV	**LLEKAA**GT**LFGNV**FS**IL**AY**AKPVIIGIILIILLLLVIRILRWLAVE	-RRRKQE----
BSV	--SINNWRR**L**QI**T**GQ**S**M**N**QA**I**TT**L**SK**L**SDLNDENLAAGIHLLQDHIV	**LLEKAA**GT**LFGNV**FS**IL**AY**AKPVIIGIILIILLLLVIRILRWLAVE	--RRRKQE----
EFV	---GDNFRK**L**QT**S**GL**S**M**N**QA**I**ST**I**AK**I**SDLNDENLAAGIHLLQEHIV	**FLGKIA**DG**IFGTT**FS**LL**TY**AKPVIIGIVVIVLLILIRILSWLAALGPRRRRREDKGE	
SloEFV	RMKKPNTQR**L**EK**V**SL**M**M**A**NS**T**AT**V**SK**L**SDLNEYLFADGLHILKDHVV	**FVSSAT**KG**IFGGI**ID**IL**TY**TKPIVILIIITILIVLIFRILKWLPNSE	--KKKEQ----

Fig. 1

Analyses of gp47 amino acid sequences

(a) The hydrobicity prediction of PFV gp47. Two hydrophobic regions had been identified in PFV gp47-N-FP and a longer region including MPER and MSD. (b) The schematic diagram and amino acid sequences of the predicted important motifs of FV gp47. The motifs had been labeled, including FP, polar region, N-HR, C-HR, MPER, MSD and ERRS motif. The corresponding amino acid sequences were indicated. The FP and MPER showed a typical i, i+3/4, i+7 pattern (boxed), which was characteristic of amphipathic α-helix. (c) The helical wheel analysis of PFV FP. The arrow pointed to the hydrophobic face (AIAMLLLV) of the FP.

helix regions at N-terminus and C-terminus respectively, and a long region of extended β-strands and loops in the middle. The results were similar to the previous analyses using a different secondary structure prediction program (Sun *et al.*, 2012).

Fusion peptide (FP) of FVs gp47

A prominent functional region of retroviral Env TM subunits was the hydrophobic N-terminal fusion peptide (FP). While the FVs gp47 were analyzed using the Kyte-Doolittle

hydrophobicity plot, the hydrophobic region had not been identified in the N-terminus (Wang and Mulligan, 1999). Until the White and Wimley hydrophobicity prediction algorithm program, Membrane Protein Explorer (MPEx), was employed, the PFV N-terminal hydrophobic FP was ultimately detected (Fig. 1a). The reason might be that the MPEx program was using "interface scale" method rather than overall average hydrophobicity per residue. However, the PFV FP was predicted to be close to the N-terminus of gp47, not at the immediate N-terminus as in some other retroviruses, such as HIV, SIV, HBV and MMTV. Also, similar results were obtained in other FVs (data not shown).

Sequence alignment indicated that the FP possessed a typical i, i+3/4 and i+7 pattern (Fig. 1b), which tended to form an amphipathic α-helix structure. The helical wheel analysis of PFV gp47 FP revealed that the peptide showed a significant hydrophobic face (AIAMLLLV) and maintained a high hydrophobicity index (H.I.) and hydrophobic moment (Fig. 1c). The characteristics of all 15 FV gp47 helical wheel analyses were summarized in Table 2. Generally, the retroviral N-terminal FP had common features, such as the existing of a high content of A, G and F and tandem repeats of the tripeptide FLG (Gallaher, 1987; Peisajovich and Shai, 2003). But these features were not observed in the FV N-terminal FP. However, it was interesting that a positively charged K/R residue was detected at the N-terminus of the potential FP in all the FVs (Fig. 1b).

There were 15 polar residues following the HIV gp41 FP. And these polar residues significantly enhanced the gp41

Table 2. The characteristics of FPs of FVs and several other retroviruses

Virus	Hydrophobicity index	Hydrophobic moment	Hydrophobic face
PFV	0.551	0.425	AIAMLLLV
SFVagm	0.532	0.421	IAMLLLV
SFVcpz	0.552	0.424	AIAMLLLV
SFVgor	0.571	0.421	AIAMLLLVGA
SFVmac	0.551	0.425	AIAMLLLV
SFVmar	0.446	0.451	AIAAII
SFVorg	0.489	0.461	IAMLLLV
SFVspm	0.399	0.452	AIILVGA
SFVsqm	0.548	0.426	VIILVGA
BFV	0.483	0.390	LMLI
BSF	0.483	0.390	LMLI
EFV	0.496	0.328	LMLIGA
FFV	0.518	0.382	IAAVLIIGAA
FSV	0487	0.352	IAAVLIIGAA
SloEFV	0.433	0.343	VVML
HIV-1	0.731	0.250	FIAFALGGL
SIV	0.892	0.249	PAGAFFAFMLVGLVGL
HBV	0.847	0.342	VGGPILLFMFL
MMTV	0.589	0.158	AALFN

fusogenic activity by model PC/PE/Cho (1:1:1) membranes fusion assay (Peisajovich and Shai, 2003). The gp47 sequence of PFV was examined and the polar segment, SDINDEN-LQQGIYLLRDH, following the FP had been identified. Similarly, the polar regions in other 14 FVs gp47 had been uncovered as well (Fig. 1b). So the broad gp47 subunit FP should include the N-terminal hydrophobic region and the following polar residues.

Membrane-spanning domain (MSD) of FVs gp47

Another prominent hydrophobic region near the C-terminus of FVs gp47 was predicted to be the membrane-spanning domain (MSD) (Fig. 1a), which plays an important role in regulating fusogenic activities of FVs and the virion particles egress (Pietschmann *et al.*, 1999, 2000). Based on the *in sillico* analysis, the MSD consisting of 23 hydrophobic amino acids (AFSLLGYLKPILIGVGVILLVILIF) in the C-terminal region of the PFV gp47 had been identified. Also, the MSDs of other FVs gp47 were predicted (Fig.1b). Although the locations of MSDs were not fixed, these sequences had mild extent of sequence conservation. Surprisingly, at position 9 of PFV MSD sequence, a conserved, positively charged K residue followed by a P residue, known as a helix breaker, could be found, which was involved in the regulation of Env fusion activity and intracellular transport (Pietschmann *et al.*, 2000). In addition, there was also a positively charged K/R near the C-terminus of MSD. The FV MSDs, except EFV, were followed by positively charged residues (K or R), which were named as the dilysine motif (lysine at -3, lysine or arginine at -4 or/and -5 relative to the C-terminus), also known as the endoplasmic reticulum retrieval signal (ERRS). Based on this motif, most FVs were observed budding predominantly at endomembranes (e.g., ER, Golgi). When the dilysine motif was missing, EFV was budded at the plasma membrane (Muñoz-Barroso *et al.*, 1999; Lindemann and Rethwilm, 2011; Sun *et al.*, 2012).

Membrane proximal ectodomain region (MPER) of FVs gp47

The segment located at the C-terminus of HIV-1 gp41 ectodomain displayed a high tendency to partition into the membrane interface. Thus, this Trp-rich fragment (LWNWFNITNWLWYIKLFIM) was named as membrane proximal ectodomain region (MPER), which was identified to be crucial for viral entry and complete membrane fusion (Muñoz-Barroso *et al.*, 1999; Salzwedel *et al.*, 1999; Garg and Blumenthal, 2008; Garg *et al.*, 2011). For PFV gp47, the prominent hydrophobic domain near the C-terminus was particularly long. Thus, it was speculated that the domain should include MSD and MPER segments. The assumption had been verified by the prediction using Interfacial

hydrophobicity scale program MPEx (Fig. 1a). The MPER segment (FLSGTAQGIFGTAFSLLGY), which had seven residues overlapping with the MSD, had been found in PFV. Interestingly, the PFV MPER, similar to the N-terminal FP, showed a typical i, i+3/4 and i+7 pattern which indicated it had a high potential to form an amphipathic α-helix structure. The similar characteristics had been revealed in other 14 FVs gp47 (Fig. 1b).

Several previous studies revealed that the membrane-destabilizing activities of the MPER segments from FVs gp47 were similar to the equivalent region from HIV-1 gp41 (Epand and Epand, 2001; Peisajovich and Shai, 2003). It was also suggested that during the fusion process of FV, the N-terminal FP and the C-terminal MPER might destabilize the target cell membrane and the viral envelope, respectively (Peisajovich and Shai, 2003).

The functional similarities between the FP and MPER segment suggested that the two regions might have commonalities in their sequences. However, it should be noted that the interactions between the target cell membrane and the FP, and between the viral envelope and the MPER had opposite topology (Epand and Epand, 2001; Peisajovich and Shai, 2003). Alignments between the PFV MPER and the "reversed" N-terminal FP sequence showed a medium similarity (Fig. 2). A higher identity between the PFV MPER and the "reversed" N-terminal FP of HIV gp41 was observed. Strikingly, the reversed-tripeptide FIG in the PFV gp47 sequence was aligned with the HIV-1 gp41 FLG tripeptide. Similar results were also obtained for the other 14 FVs (supplementary Fig. 2). In addition, being different from the FV

```
PFV MPER          FLSGTAQGIFGT-AFSLLGY  19
Reversed PFV FP   -IQSLTQ-VAGTLAYGMSRL  18

                  :.. :* : ** *:.:

PFV MPER          FLSGTAQGIFGTAFSLLGY-  19
Reversed HIV FP   -TSG-AAGLFGLFLAGIGVA  18

                  ** * *:** :: :*
```

Fig. 2

The sequence alignment between PFV MPER and N-FP of PFV and HIV

PFV MPER showed a higher similarity with reversed HIV N-FP, and a "FIG" tripeptide corresponding to HIV "FLG" had been identified.

FPs, the MPERs of FVs gp47 had a high content of A, G and F residues that were known to be important for the fusogenic activity (Epand and Epand, 2001).

N-terminal heptad repeat (N-HR) and C-terminal heptad repeat (C-HR) of FVs gp47

In several viral envelope-fusion domains, the sections of α-helical coiled coil following the FP had a high α-helical propensity and a 4–3 heptad repeat of hydrophobic residues. In PFV, the similar peptide fragments named as N-HR and C-HR were predicted in the ectodomain of gp47. The predicted N-HR peptide segment, prior to the central

Table 3. The predicted N-HR and C-HR of FVs gp47

Virus	N-HR d a d a d a d	C-HR
PFV	TWLQQQLQKSDDEMKVIKRIARSLV	IEVTSSGESIKEQIERAKAELLRLDIHE
SFVagm	NWIKEQLQKTEDEMKIIRRTAKSLV	IEVTSTQESIKDQIERAKSQLLRLDIHE
SFVcpz	AWLQQQLQKSDDEMKVIKRIAKSLV	IEVTSSGESIKDQIERAKAELLRLDIHE
SFVgor	SWLQTQLQKSDDEMKVIKRTARSLV	IEVTSSGESIKDQLERAKAELLRLDIHE
SFVmac	DWIQQQLQKTDDEMKLIRRTARSLV	IEVTSTQENIKDQIERAKAQLLRLDIHE
SFVmar	RWIQDQLGLDEADMKILRRTARALI	IQVTSTWESIKDQVEKSQTELLRLDIHE
SFVspm	QWIKTQLQLNDEDMKVLRRTARALV	VKITSTWETIKDQIDRAQAELLRLDLHE
SFVsqm	SWIQDQLKLSDEDMKILRRASRALV	VQITSTWESIKGQVEQAQAELLRLDLHE
EFV	TWIQQQLQADEALMNVIRRTARSMT	IEITSTWENIKDQIKRSEAELLRLDLHE
SFVorg	DWLQQQLQQPTDHMKIIKRTARSLV	IEVTSTQENIKDQIERAKAELLRLDIHE
BFV	KWIQEELKYTDEVMNVIRRTARSIT	IKVTSTWESIKDQIHRSEQELLRLDLHE
BSV	KWIQEELKYTDEVMNVIRRTARSIT	IKVTSTWESIKDQIHRSEQELLRLDLHE
FFV	SWIQEELGVSDNIMKVIRKTARCIP	IEITSSWETIKEQVARAKAELLRLDLHE
FSV	SWIQEELGVSDNIMKVIRKTARCIP	IEITSSWETIKEQVARAKAELLRLDLHE
SloEFV	TWINEGLNIPHHGMQIVKRMSCSNI	VKATDTWASIEEQIEDTKSDLLRLELHK

The N-HR of BFV, BSV, FFV, FSV, SFVorg and SloEFV were identified manually. As with other FVs, the positions 'a' and 'd' of BFV, BSV, FFV, FSV, SFVorg and SloEFV sequences are mostly hydrophobic or neutral amino acids (in gray).
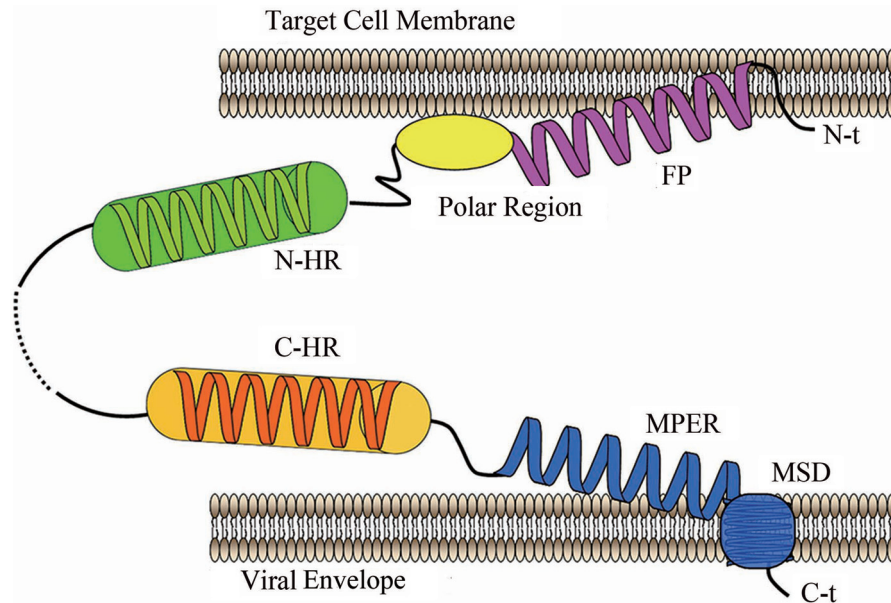
Fig. 3

The diagram of interaction between the gp47 fusion domains and the target membranes (drawn by Science Slides Plug-In)
The N-terminus of viral FP (in purple helix) inserted into the target cell bilayer, and the following polar region (in yellow oval) lay near the target membrane surface to promote the fusogenic activity. The C-terminus of MPER (in blue helix, a part of the MSD in most FVs) contacted with the viral envelope along with the MSD (in blue tubular helix), which was inserting into the viral envelope. The N-HR and the C-HR were symbolized by the green and orange helices, respectively. The central β-strands and loops region were not shown in this diagram.

strands and loops region, was 25 amino acids in length (TWLQQQLQKSDDEMKVIKRIARSLV) and fitted in with the characteristic of 4–3 hydrophobic repeat region. The C-HR peptide was 28 amino acids in length (IEVTSSGE-SIKEQIERAKAELLRLDIHE) and located prior to the MPER segment (Fig. 1b). According to the computer predictions, the C-HR had been found in all 15 FVs gp47, but the N-HRs were only detected in some of FVs (Table 3). This might confirm the assumption that the C-HR showed stronger effect than the N-HR, like in HIV gp41 (Herskowitz, 1987; Chan *et al.*, 1997; Yueyong *et al.*, 2006).

Discussion

While HIV FP begins with gp41 N-terminus, FVs has an internal FP that is close to the gp47 N-terminus but not immediately N-terminal. There is a positively charged K/R at the front of the FP, which might be involved in the regulation of fusion activity, just like the K/R in MSD. In HIV gp41, FP could insert into membranes and promote membrane negative curvature. As it had been reported that FP could minimally assemble into trimer by gp41 trimeric coiled coil and six-helix bundle (Sackett *et al.*, 2010), it was suggested that the FP trimer could induce fusion better than its

monomer form and had a strong correlation with the depth of insertion (Yang and Weliky, 2003; Qiang and Weliky, 2008; Garg *et al.,* 2011). While different from HIV FP and some other retroviral FPs, FV FP promoted the fusogenic activity apparently by a mechanism that was not simply dependent on changes in membrane curvature (Martin *et al.*, 1999; Epand and Epand, 2001; Epand, 2003). Because there were several amino acids preceeding the FV FPs, it had been suspected that these amino acids could accelerate the rate of membrane fusion by some other means, although the positive curvature would be promoted (Epand and Epand, 2001).

For HIV, except the insertion of FP into the target cell membrane, the polar residues following FP might lie near the surface of the membrane and destabilize membrane by forming a correct oligomerization (Peisajovich *et al.*, 2000; Peisajovich and Shai, 2003). In our study, it was found that although having a low overall hydrophobicity index, the FV FPs presented a high propensity to form an amphipathic α-helix structure, which showed a high hydrophobic moment and maintained a high hydrophobicity index at the hydrophobic face. Also, similar to HIV, several polar residues following the FPs were found in FVs. So a hypothesis was proposed that the FV polar region, which followed the FV FPs that inserted into and destabilized the target cell membranes, might further stimulate the fusogenic activities

through touching the surface of the membranes and promoting their self-association. Based on this functional relationship, we hypothesized that in FVs the FP and the following polar region might be working together to make up a broad gp47 fusion peptide motif.

The MPER of HIV gp41 was critical for Env incorporation into the virion as well as functionally for its ability to perturb membrane, explaining its high level of conservation (Vishwanathan and Hunter, 2008; Garg *et al.,* 2011). Mutation of more than one Trp in HIV Trp-rich region severely affected gp41-mediated fusion and infectivity (Salzwedel *et al.*, 1999; Suárez *et al.*, 2000; Garg and Blumenthal, 2008). But the Trp-rich region was not observed in FV MPER sequences. However, the MPERs of the FVs gp47 displayed a high content of A, G and F, and had a conserved "reversed" FIG or FLG tripeptide, which showed a high probability to localize at the bilayer interface. Furthermore, a synthetic peptide corresponding to PFV MPER had been demonstrated to destabilize membrane and induce fusion of model PC/PE/ Cho (1:1:1) membranes (Epand and Epand, 2001). Thus, it was suggested that the MPER could insert into the viral envelope, like the N-terminal FP that could insert into the cell membrane (Peisajovich and Y, 2002). But the mechanisms of the MPER and FP in FVs might not be the same, as only slight similarities exist between the two motifs. For example, the FIG or FLG tripeptide had not been identified in the FV FPs. For most FVs, the MPER was identified to overlap with the MSD motif according to our predictions, which meant that the amino acids at the C-terminus of MPER were also part of the MSD. Thus, it would be better to illustrate that the FVs fusion process needed several amino acids of MPER inserting and the viral envelope disturbing.

For some retroviruses, including FV, the formation of the α-helical coiled coils had been postulated to relocate the N-terminal FP and C-terminal MPER close to the target cell endosomal membrane and viral lipid envelope, respectively (Peisajovich and Shai, 2002; Peisajovich and Shai, 2003). Therefore, the 15 FV heptad repeats that could form the coiled coils had been analyzed. At first, both of the N-HR and C-HR were discovered, which respectively derived from the N- and C-terminal regions of the PFV gp47. However, only the C-HRs had been clearly detected in the other 14 FVs gp47. When another program had been used, although the score was less than the C-HRs, the N-HRs had been identified in some of the FVs gp47. Undeniably, the N-HRs of BFV, BSV, FFV, FSV, SFVorg and SloEFV were not yet detected. We speculated, however, that the coiled coils at the N-terminus still existed in these FVs gp47. Firstly, it was no doubt that the helix structures, which would be necessary for coiled-coil forming, had been found in all of the FVs gp47 (sFig. 1). In addition, coiled coil regions consisted of seven-residue repeats of amino acids in a sequence periodicity (a b c d e f g), in which the side-chains of amino acids in positions 'a', 'd' were predominantly bulky

and hydrophobic or neutral (Mclachlan and Karn, 1983; Yueyong *et al.*, 2006). Thus, the N-HRs of BFV, BSV, FFV, FSV, SFVorg and SloEFV had been identified manually (Table 3). Similar sequence features had been found in these N-HRs, for example, in the positions 'a' and 'd' were hydrophobic or neutral amino acids. Thus, we could draw a conclusion that the N-HRs did exist in these FVs.

Taken together, our results illustrated that FVs gp47 had many conserved motifs that were involved in the process of membrane fusion (Fig. 3). The C-terminal MSD might insert into the viral envelope, and the MPER acted on the viral envelope mirroring the activity of the N-terminal FP on the cell membrane. The FV polar region laid on the surface of the cell membrane would stimulate the fusogenic activity. The HRs, which were linked by the β-strands and loops region, would assist the N-terminal FP and the C-terminal MPER during membrane merging. It had been proposed that the HRs of influenza HA2, HIV-1/SIV gp41, Ebola virus GP2 and Simian virus 5 F1 were able to form six-helix bundles in a post-fusion form (Chan *et al.*, 1997; Malashkevich, 1999; Yueyong *et al.*, 2006; Wang *et al.*, 2012). For FVs gp47, a six-helix bundle could also been formed, in which three N-HRs formed an interior N-helix trimer, and three C-HRs inserted into the highly conserved, hydrophobic cavities on the surface. The packing of the C-HR against the N-HR grooves, which were considered as fusion core, would bring the viral envelope and the cell membrane close to one another. Taken together, the fusion domains of FV TM subunits gp47 played extremely important roles in fusion between viral envelope and target cell membrane. The results obtained in this work would stimulate and guide the study of the three dimensional molecular structures of FV glycoprotein and help better understanding the virus Env-mediated cell fusion mechanism.

References

Aris K, Gifford RJ, Michael T, Gilbert MTP, Pybus OG (2009): Macroevolution of complex retroviruses. Science 325, 1512. http://dx.doi.org/10.1126/science.1174149

Berger B, Singh M (1997): An iterative method for improved protein structural motif recognition. J. Comp. Biol. 4, 261–273. http://dx.doi.org/10.1089/cmb.1997.4.261

Chan DC, Fass D, Berger JM, Kim PS (1997): Core structure of gp41 from the HIV envelope glycoprotein. Cell 89, 263–273. http://dx.doi.org/10.1016/S0092-8674(00)80205-6

Dirk L, Axel R (2011): Foamy Virus Biology and Its Application for Vector Development. Viruses 3, 561–585. http://dx.doi.org/10.3390/v3050561

Duda A, Stange A, Lüftenegger D, Stanke N, Westphal D, Pietschmann T, Eastman SW, Linial ML, Rethwilm A, Lindemann, D (2004): Prototype foamy virus envelope glycoprotein leader peptide processing is mediated by a furin-like cellular protease, but cleavage is not essential for viral infectivity. J. Virol. 78, 13865–13870. http://dx.doi.org/10.1128/JVI.78.24.13865-13870.2004

Epand RM, Epand RF (2001): Factors Contributing to the Fusogenic Potency of Foamy Virus. Biochem. Biophys. Res. Commun. 284, 870–874. http://dx.doi.org/10.1006/bbrc.2001.5060

Epand RM (2003): Fusion peptides and the mechanism of viral fusion. Biochim. Biophys. Acta 1614, 116–121. http://dx.doi.org/10.1016/S0005-2736(03)00169-X

Gallaher WR (1987): Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell 50, 327–328. http://dx.doi.org/10.1016/0092-8674(87)90485-5

Garg H, Blumenthal R (2008): Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. Cell. Molec. Life Sci. 65, 3134–3144. http://dx.doi.org/10.1007/s00018-008-8147-6

Garg H, Viard M, Jacobs A, Blumenthal R (2011): Targeting HIV-1 gp41-induced fusion and pathogenesis for anti-viral therapy. Curr. Top. Med. Chem. 11, 2947–2958. http://dx.doi.org/10.2174/156802611798808479

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (1999): Protein identification and analysis tools on the ExPASy server. Methods Mol. Biol. 112, 531–552.

Gautier R, Douguet D, Antonny B, Drin G (2008): HELIQUEST: a web server to screen sequences with specific α-helical properties. Bioinformatics 24, 2101–2102. http://dx.doi.org/10.1093/bioinformatics/btn392

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010): A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res. 38, W695–W699. http://dx.doi.org/10.1093/nar/gkq313

Herskowitz I (1987): Functional inactivation of genes by dominant negative mutations. Nature 329, 219–222. http://dx.doi.org/10.1038/329219a0

Linial, M (2000): Why aren't foamy viruses pathogenic? Trends Microbiol. 8, 284–289. http://dx.doi.org/10.1016/S0966-842X(00)01763-7

Lu M, Blacklow SC, Kim PS (1995): A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nat. Struct. Biol. 2, 1075–1082. http://dx.doi.org/10.1038/nsb1295-1075

Lupas A, Van Dyke M, Stock J (1991): Predicting coiled coils from protein sequences. Science 252, 1162–1164. http://dx.doi.org/10.1126/science.252.5009.1162

Malashkevich VN (1999): Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. Proc. Natl. Acad. Sci. USA 96, 2662–2667. http://dx.doi.org/10.1073/pnas.96.6.2662

Martin I, Ruysschaert JM, Epand RM (1999): Role of the N-terminal peptides of viral envelope proteins in membrane fusion. Adv. Drug Deliv. Rev. 38, 233–255. http://dx.doi.org/10.1016/S0169-409X(99)00031-9

Mclachlan AD, Karn J (1983): Periodic features in the amino acid sequence of nematode myosin rod. J. Mol. Biol. 164, 605–626. http://dx.doi.org/10.1016/0022-2836(83)90053-0

Mu-oz-Barroso I, Salzwedel K, Hunter E, Blumenthal R (1999): Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion. J. Virol. 73, 6089–6092.

Peisajovich S, Epand RM, Shai Y, Epand R (2000): The polar region consecutive to the HIV fusion peptide participates in membrane fusion. Biochemistry 39, 1826–1833. http://dx.doi.org/10.1021/bi991887i

Peisajovich SG, Shai Y (2003): Viral fusion proteins: multiple regions contribute to membrane fusion. Biochim. Biophys. Acta 1614, 122–129. http://dx.doi.org/10.1016/S0005-2736(03)00170-6

Peisajovich SG, Shai Y (2002): High similarity between reverse-oriented sequences from HIV and foamy virus envelope glycoproteins. AIDS Res. Hum. Retroviruses 18, 309-312. http://dx.doi.org/10.1089/088922202753472883

Pietschmann T, Heinkelein M, Heldmann M, Zentgraf H, Rethwilm A, Lindemann D (1999): Foamy virus capsids require the cognate envelope protein for particle export. J. Virol. 73, 2613–2621.

Pietschmann T, Zentgraf H, Rethwilm A, Lindemann D (2000): An evolutionarily conserved positively charged amino acid in the putative membrane-spanning domain of the foamy virus envelope protein controls fusion activity. J. Virol. 74, 4474–4482. http://dx.doi.org/10.1128/JVI.74.10.4474-4482.2000

Qiang W, Weliky DP (2008): HIV fusion peptide and its cross-linked oligomers: efficient Syntheses, significance of the trimer in fusion activity, correlation of β strand conformation with membrane cholesterol, and proximity to lipid headgroups. Biochemistry 48, 289–301. http://dx.doi.org/10.1021/bi8015668

Sackett K, Nethercott MJ, Epand RF, Epand RM, Kindra DR, Shai Y, Weliky DP (2010): Comparative analysis of membrane-associated fusion peptide secondary structure and lipid mixing function of HIV gp41 constructs that model the early pre-hairpin intermediate and final hairpin conformations. J. Mol. Biol. 397, 301–315. http://dx.doi.org/10.1016/j.jmb.2010.01.018

Salzwedel K, West JT, Hunter E (1999): A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. J. Virol. 73, 2469–2480.

Snider C, Jayasinghe S, Hristova K, White SH (2009): MPEx: a tool for exploring membrane proteins. Protein Sci. 18, 2624–2628. http://dx.doi.org/10.1002/pro.256

Suárez T, Gallaher WR, Agirre A, Go-i FM, Nieva JL (2000): Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fu-

sion. J. Virol. 74, 8038–8047. http://dx.doi.org/10.1128/JVI.74.17.8038-8047.2000

Sun Y, Wen DD, Liu QM, Yi XF, Wang TT, Wei LL, Li Z, Liu WH, He X-H (2012): Comparative analysis of the envelope glycoproteins of foamy viruses. Acta Virol. 56, 283–291. http://dx.doi.org/10.4149/av_2012_04_283

Tusnady GE, Simon I (1998): Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J. Mol. Biol. 283, 489–506. http://dx.doi.org/10.1006/jmbi.1998.2107

Vishwanathan SA, Hunter E (2008): Importance of the membrane-perturbing properties of the membrane-proximal external region of human immunodeficiency virus type 1 gp41 to viral fusion. J. Virol. 82, 5118–5126. http://dx.doi.org/10.1128/JVI.00305-08

Wang G, Mulligan MJ (1999): Comparative sequence analysis and predictions for the envelope glycoproteins of foamy viruses. J. Gen. Virol. 80, 245–254. http://dx.doi.org/10.1099/0022-1317-80-1-245

Wang X, Xiong W, Ma X, Wei M, Chen Y, Lu L, Debnath AK, Jiang S, Pan C (2012): The conserved residue Arg46 in the N-terminal heptad repeat domain of HIV-1 gp41 is critical for viral fusion and entry. PLoS One 7, e44874. http://dx.doi.org/10.1371/journal.pone.0044874

Yang J, Weliky DP (2003): Solid-state nuclear magnetic resonance evidence for parallel and antiparallel strand arrangements in the membrane-associated HIV-1 fusion peptide. Biochemistry 42, 11879–11890. http://dx.doi.org/10.1021/bi0348157

Yueyong L, Yanhui X, Zhiyong L, Jieqing Z, Xuebo H, Gao GF, Bingsheng Q, Zihe R, Po T (2006): Structural characterization of Mumps virus fusion protein core. Biochem. Biophys. Res. Commun. 348, 916–922. http://dx.doi.org/10.1016/j.bbrc.2006.07.168

Supplementary information

Structure of transmembrane subunits gp47 of the foamy virus envelope glycoproteins

M. WANG[1], H. ZHANG[1], Q.-M. LIU[1], Y. SUN[1*], Z. LI[1*], W.-H. LIU[2], X.-H. HE[2], J. SONG[1], Y.-X. WANG[1]

[1]College of Life Sciences, Shaanxi Normal University, Xi'an 710062, P. R. China; [2]Pathogenic Organism and Infectious Diseases Research Institute, School of Basic Medical Sciences, Wuhan University, Wuhan 430071, P. R. China
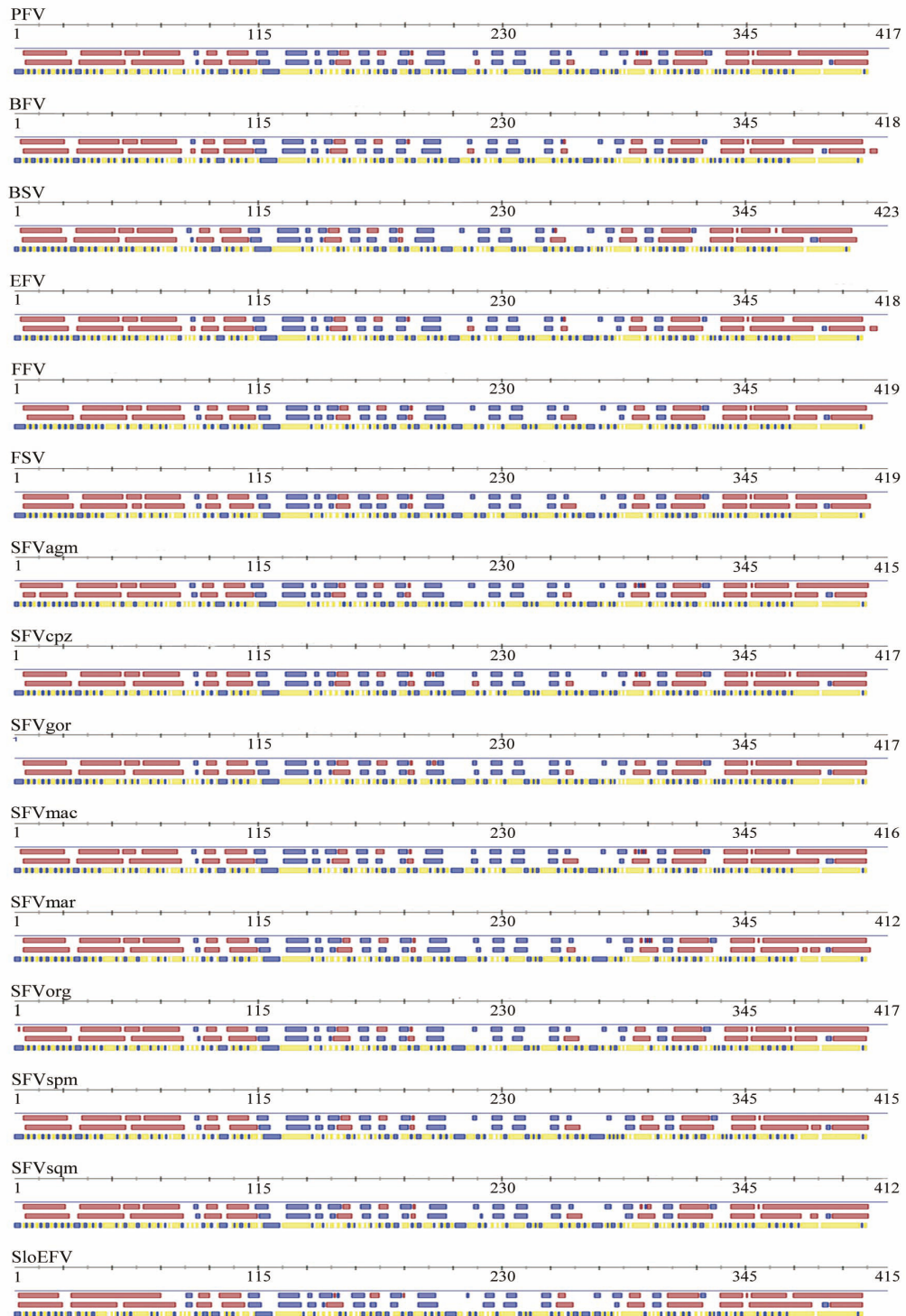
Supplementary Fig. 1

Prediction for secondary structure and solvent accessibility of each FV gp47

The secondary structure had been analyzed using Predictprotein program, RePROF (the first line) and PROFsec (the second line), respectively (the helix in red; the strand in blue; the loop in blank). The similar results were obtained from two prediction programs. The results showed that N-terminus and C-terminus were helix regions and the middle is a long region of strands and loops. The solvent accessibility of gp47 had also been displayed (the third line, the buried in yellow; the exposed in blue; the intermediate in blank). A region full of buried residues had been identified at the N-terminus as MSD reigon.

```
SloEFV MPER  FVSSATKGIFGGIIDILTY-        SFVcpz MPER  FLSGAAHGIFGTAFSLLGY-
Rev HIV FP   --TSGAAGLFGLFLAGIGVA        Rev HIV FP   -TSGAA-GLFGLFLAGIGVA
                :*.: *:** ::  :                       **** *:**  :: :*

SFVgor MPER  FLTGAAQGLFGTAFSILGY-        SFVorg MPER  FLSDAAQGIFGTAFGILGY-
Rev HIV FP   --TSGAAGLFGLFLAGIGVA        Rev HIV FP   --TSGAAGLFGLFLAGIGVA
                *..* ****  :: :*                      :..* *:**  :. :*

SFVagm MPER  VLSNTAQGIFGTTVSILSY-        SFVmac MPER  FLSNTAQGIFGSAVS-LLFY
Rev HIV FP   --TSGAAGLFGLFLAGIGVA        Rev HIV FP   --TSGAAGLFGLFLAGIGVA
                :. * *:**  :: :.                      :. * *:**  :: : .

SFVmar MPER  FLSGTFGGLFGTLGYIKPI---      SFVspm MPER  FLSGTFGGLFGTLGYIKPI---
Rev HIV FP   -TSGAAG-LFG-L-FLAGIGVA      Rev HIV FP   -TSGAAG-LFG-L-FLAGIGVA
                **: * *** * ::   *                    **: * *** * ::   *

SFVsqm MPER  FLGGTAGSIFGIF--GYLKPI       FFV MPER     FIKDTAGGIFGTAFSFLGY-
Rev HIV FP   -TSGAAG-LFGLFLAG-IGVA       Rev HIV FP   --TSGAAGLFGLFLAGIGVA
                .*:** :**:*  * :                      .. *.*:**  :: :*

BFV MPER     LLEKAAGTLFGNVFSILAY-        EFV MPER     FLGKIADGIFGTTFSLLTY-
Rev HIV FP   -TSGAAG-LFGLFLAGIGVA        Rev HIV FP   TSG--AAGLFGLFLAGIGVA
                . *** *** .:: :.                      *  * *:**  :: :

BSV MPER     LLEKAAGTLFGNVFSILAY-        FSV MPER     FIKDTAGGIFGTAFSFLGY-
Rev HIV FP   -TSGAAG-LFGLFLAGIGVA        Rev HIV FP   --TSGAAGLFGLFLAGIGVA
                . *** *** .:: :.                      .. *.*:**  :: :*
```

Supplementary Fig. 2

The sequence alignment between MPER of different foamy viruses and HIV N-FP
All the 14 MPERs showed similarities with reversed HIV N-FP, and the "FIG" tripeptide had been identified in most of the MPERs. (Rev = Reversed).