Journal of Animal and Veterinary Advances 11 (21): 3957-3968, 2012
ISSN: 1680-5593

The Characteristics of Context-Dependent Codon Usage Bias for Cleavage Sites in the Polyprotein of Foot and Mouth Disease Virus

Jun-Lin Liu, Xiao-Xia Ma, Shi-En Che, Ji-Ying Xu, Yi-Xia Chen, Zhuo Li, Yu-Ping Fen, Ju-Tian Yang, Hong Chen and Sheng-Dong Huo
College of Life Science and Engineering, Northwest University for Nationalities, 730030 Lanzhou, P.R. China

Abstract: The results of analysis of context-dependent codon bias designed to identify the relationship between codon bias model for the context surrounding the cleavage site and the model for the corresponding cleavage site, applied to genome data form 66 isolates of Foot and Mouth Disease Virus (FMDV) representing all seven serotypes were reported. The cleavage sites in the polyprotein were identified from all parts of the FMDV genome and the similarity of Codon Bias Model for regions flanking the cleavage sites was estimated. Although, the two amino acid residues which was comprised of the cleavage site were not conserved, one residue always tended to be conserved, suggesting that the conserved residue was essential for the cleavage process. In addition, although the similarity of Codon Bias Model for some positions in the most target contexts was ensured, the scale of similarity of Codon Bias Model was larger in the regions flanking the 2A/2B, 2C/3A, 3A/3B, 3B/3C and 3C/3D sites than that of the rest. Furthermore, correlation analysis between the Codon Bias Model for each position in the target context surrounding the cleavage site and that of the corresponding site was performed it was concluded that the significant relationship existed in the VP4/VP2, 2A/2B, 2B/2C and 3A/3B sites, respectively. The interesting phenomena probably suggested that context-dependent codon bias surrounding some cleavage sites was existence in FMDV and the Codon Bias Model for these regions played potential roles in shaping specific characteristics of these corresponding sites at the level of transcription or translation.

Key words: Foot and mouth disease virus, the cleavage site, context-dependent codon bias, Codon Bias Model, China

INTRODUCTION

Foot and Mouth Disease Virus (FMDV) which contains one single-stranded positive-sense RNA in icosahedral particles without envelop was classified as a number of the *Aphthovirus* genus of the family Picornaviridae. The virus existed in the form of seven different serotypes: O, A, C, Asia 1, South African Territories 1 (SAT 1), SAT 2 and SAT 3 but many subtypes had evolved within each serotype. This virus possessed the single stranded positive-sense RNA genome which was used as an mRNA and encodes a viral polyprotein. This polyprotein was observed neither in infected cells nor translation reactions *in vitro* because of the primary cleavage. The polyprotein of the primary cleavages was mediated by the virus encoded two proteinases (L^{pro} and $3C^{pro}$) and a short oligopeptide sequence (2A). For L^{pro} it cleaved the cleavage site between L^{pro} and VP4 to create it's own C-terminus (Cao *et al*., 1995; Medina *et al*., 1993). For $3C^{pro}$, it cleaved the 2C/3A site (Bablanian and Grubman, 1993; Clarke and Sangar, 1988; Vakharia *et al*., 1987). For 2A it performed the primary cleavage between it's own C-terminus and N-terminus of protein 2B by a translational mechanism rather than proteolytic (Vakharia *et al*., 1987; Donnelly *et al*., 1997; Ryan *et al*., 1991). For the 5' and 3' untranslated region (5' and 3' UTR) of FMDV they played an important role in the infectivity and viral RNA replication. And Lopez de Quinto *et al*. (2002) found that the FMDV 3'UTR sequence could stimulate the activity of the Internal Ribosome Entry Site (IRES) and suggested that individual signals in the 3'UTR carry out stimulation of FMDV translation (Lopez de Quinto *et al*., 2002). This interesting result probably indicated that for FMDV RNA genome a potential interaction of intragenic region influenced translation levels of FMDV various products. However, the biological functions of the viral products were mainly focused by many reports (Grubman and Baxt,

Corresponding Author: Shi-En Che, College of Life Science and Engineering, Northwest University for Nationalities, 730030 Lanzhou, P.R. China

2004; Grubman *et al*., 2008; Joern, 2009). During the analysis of evolution of FMDV genome, recombination, host range and infectivity were mainly focused by some reports (Carrillo *et al*., 2005; Cooke and Westover, 2008; Jackson *et al*., 2007; Lewis-Rogers *et al*., 2008; Zhou *et al*., 2011, 2010a, b).

It was well known that FMDV genome was far from static and possessed an ability to avoid extinction by mutation (Pariente *et al*., 2003). It was astonished that the great variety of FMDV genome never impaired this virus biology. Carrillo *et al*. (2005) indicated that 58% of the amino aicds encoded by FMDV genomes were invariant and these residues had a strong correlation with biological function of virus (Carrillo *et al*., 2005). There were some active evidences that indicated:

- A short 2A peptide along with the N-terminal praline of 2B protein possessed a property of cleavage of the products flanking this region (Ryan *et al*., 1991; Ryan and Drew, 1994) and influenced the amount of products flanking 2A region (Donnelly *et al*., 2011)
- The 5' and 3' UTRs flanking Open Reading Frame (ORF) regulated the translational efficiency (Lopez de Quinto *et al*., 2002)
- A mechanism about the alternative two AUGs in FMDV ORF might be mediated by the context flanked by the two AUGs (Belsham, 1992; Zhou *et al*., 2010b)

It was true that coding regions coding for mature products existed without overlapping region in FMDV ORF. Did upstream coding sequence affect downstream one at the translation efficiency? However, there were little information about the relationship between synonymous codon usage pattern and amino acid encoded by FMDV RNA sequences. It needed to analyze whether this phenomenon was strongly correlated with Codon Usage Bias (CUB) or not. It had also been reported that synonymous codons were not chosen equally and randomly in genomes or between genomes (Lloyd and Sharp, 1992). There had been many interest in the evolution of codon usage surrounding initiation and termination codons, the base composition, codon usage pattern and amino acids usage exhibited significant deviations from a random distribution this was accentuated in highly expressed genes (Alff-Steinberger and Epstein, 1994; Brown *et al*., 1993, 1994; Miyasaka, 1999; Ohno *et al*., 2001; Sharp and Bulmer, 1988). Hooper and Berg (2000) found that synonymous codon usage variation varied at different site along a coding sequence.

In this study, researchers focused on the similarity of Codon Bias Model for the target region flanking the cleavage site in FMDV ORF and evaluated the correlation between the scale of Codon Bias Model for the context surrounding the cleavage sites in FMDV ORF and that of the corresponding cleavage site, depending on the Relative Synonymous Codon Usage (RSCU) and CUB value.

MATERIALS AND METHODS

Sequence data and analysis: The 66 complete RNA sequences of FMDV were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/) including Asia 1, n = 10 sequences; A, n = 17; C, n = 7; O, n = 19; SAT 1, n = 5; SAT 2, n = 4; SAT 3, n = 4 listed in supplementary Table 1. In order to ensure the cleavage sites in the polyprotein of FMDV, multiple sequence alignments were performed with the Clustal W (1.7) Method of DNAstar Software (7.0) for Windows with the minor manual editing.

The calculation of the relative synonymous codon usage: To investigate the characteristics of synonymous codon usage without the confounding influence of amino acid composition among different sequences, the Relative Synonymous Codon Usage (RSCU) values among different codons in the ORF of each strain was calculated. The RSCU value of the ith codon for the jth amino acid was calculated as earlier described (Sharp *et al*., 1986). RSCU:

$$RSCU = \frac{g_{ij}}{\sum_{j}^{n_i} g_{ij}} n_i$$

where, g_{ij} is the observed number of the ith codon for jth amino acid which has n_i type of the synonymous codons. The codons with RSCU values >1.0 have positive CUB while the values <1.0 have relative negative CUB. When RSCU value is equal to 1.0 it means that this codon is chosen equally and randomly.

Calculation of codon usage bias: To calculate CUB, researchers provided that statistically equal and random usage of all available synonymous codons was the neutral point (RSCU0 = 1.00) for the development of serotypes-specific codon usage (Zhou *et al*., 2010a, b) CUB:

$$CUB = \frac{\sum_{i=1}^{n} (RSCU_{ij} - RSCU_0)}{n}$$

Table 1: Genomes examined, serotype, isolate and accession numbers

Sequence numbers	Serotype	Isolate / Country (year)	Accession No.
1	FMDV-Asia 1	Cell culture/India	AY304994
2	FMDV-Asia 1	IND139-02 (cattle)/India (2002)	DQ989322
3	FMDV-Asia 1	IND 491/97; WBN117/85/India (1985)	AY687334
4	FMDV-Asia 1	IND 321/01/India (2001)	AY687333
5	FMDV-Asia 1	Asia1/WHN/CHA/06(Pig)/China (2006)	FJ906802
6	FMDV-Asia 1	Aisa1/Vietnam/QuangTri/2007/VietNam (2007)	GQ452295
7	FMDV-Asia 1	Aisa1/HNK/CHA/05/China:Hong Kong (2005)	EF149010
8	FMDV-Asia 1	Asia 1 Leb83 iso28/Lebanon (1983)	AY593800
9	FMDV-Asia 1	Asia 1-2isr13-63 iso6/Israel (1963)	AY593796
10	FMDV-Asia 1	Asia 1-1pak iso3/Pakistan (1954)	AY593795
11	FMDV-A	Epithelium (cattle)/Pakistan (2006)	EF494487
12	FMDV-A	AIRN2005_WRL(cattle)/Turkey (2005)	EF494486
13	FMDV-A	A10holland iso82/Netherlands (1942)	AY593751
14	FMDV-A	Acanefa iso48/Argentina (1961)	AY593789
15	FMDV-A	A24 argentina iso9/Argentina (1965)	AY593767
16	FMDV-A	A26arg iso74/Argentina (1966)	AY593770
17	FMDV-A	A argentina 2000 iso104/Argentina	AY593782
18	FMDV-A	Abage iso63/Brazil (1977)	AY593787
19	FMDV-A	A brazil iso67/Brazil (1958)	AY593788
20	FMDV-A	A venceslau iso70/Brazil (1979)	AY593803
21	FMDV-A	A 13brazil iso75/Brazil (1958)	AY593753
22	FMDV-A	A 16belem iso80/Brazil (1959)	AY593756
23	FMDV-A	A 17 Aguarulbos iso83/ Brazil (1967)	AY593757
24	FMDV-A	A 24cruzeiro iso71/Brazil (1955)	AY593768
25	FMDV-A	A sabana iso68/Colombia (1985)	AY593794
26	FMDV-A	A 27columbia iso78/Colombia (1967)	AY593771
27	FMDV-A	A 18zulia iso40/Venezuela (1967)	AY593758
28	FMDV-C	Cell culture/Spain	FJ824812
29	FMDV-C	Cell culture/Spain	AM409325
30	FMDV-C	Cell culture/Spain	DQ409191
31	FMDV-C	C 5arg iso60/Argentina (1969)	AY593809
32	FMDV-C	C wald iso32/United Kingdom (1970)	AY593810
33	FMDV-C	C 1ober iso88/Germany (1960)	AY593805
34	FMDV-C	C 3ind iso19/Brazil (1971)	AY593806
35	FMDV-O	Infected premise number 1378 (sheep)/United Kingdom (2001)	EF552697
36	FMDV-O	Vesicle from cloven-hoofed animal/China (2001)	EU400597
37	FMDV-O	UAE 7/99/United Arab Emirates	EU140964
38	FMDV-O	Chu-Pei/China: Taiwan	AF026168
39	FMDV-O	Swine/China: Taiwan (1997)	NC_004004
40	FMDV-O	SKR/2000/South Korea (2000)	AJ539139
41	FMDV-O	O 1campos94 iso94/Argentina	AY593819
42	FMDV-O	O taiwan97 iso106/112/China: Taiwan	AY593835
43	FMDV-O	O Penghu/China: Taiwan (1999)	AY593833
44	FMDV-O	Cattle epithelial blister/China (1958)	AF511039
45	FMDV-O	WFL/China	EF175732
46	FMDV-O	Lz/China	DQ248888
47	FMDV-O	O1Campos/Brazil (1958)	AJ320488
48	FMDV-O	FRA/1/2001 (bovine)/France (2001)	AJ633821
49	FMDV-O	O/SKR/2002 (pig)/South Korea	AH012984
50	FMDV-O	Cattle/Japan (2000)	AB079061
51	FMDV-O	O1 Geshure/Israel	AF189157
52	FMDV-O	HKN/2002/China	AY317098
53	FMDV-O	O/SKR/2000/(cattle)/South Korea	AH012985
54	FMDV-SAT 1	SAT1-1bech iso30/ Botswana (1970)	AY593838
55	FMDV-SAT 1	SAT1-3swa iso14/Namibia (1949)	AY593840
56	FMDV-SAT 1	SAT1-4srhod iso24/Zimbabwe (1958)	AY593841
57	FMDV-SAT 1	SAT1-5sa iso13/South Africa (1961)	AY593842
58	FMDV-SAT 1	SAT1-6swa iso16/Namibia (1940)	AY593843
59	FMDV-SAT 2	ZIM/7/83 (bovine)/Botswana (1983)	AF540910
60	FMDV-SAT 2	SAT2-1rhod iso26/Zimbabwe (1948)	AY593847
61	FMDV-SAT 2	SAT2-2 iso25/unknown (1967)	AY593848
62	FMDV-SAT 2	SAT2-3kenya-21/Kenya (1960)	AY593849
63	FMDV-SAT 3	SAT3-2sa iso27/South Africa (1959)	AY593850
64	FMDV-SAT 3	SAT3-3bech iso29/Botswana (1961)	AY593851
65	FMDV-SAT 3	SAT3-3kenya iso22/Kenya (1960)	AY593852
66	FMDV-SAT 3	SAT3-4bech iso23/Botswana (1965)	AY593853

More simply, CUB is the average value of difference between $RSCU_{ij}$ and $RSCU_0$ at each position of the target context. $RSCU_{ij}$ means that the value of one synonymous codon usage corresponding to a particular amino acid at one position of the context surrounding the cleavage site in FMDV ORF; n represents all codons appearing in this position. When all RSCU values according to a particular position in the target context are $RSCU_0$, CUB is equal to zero. It means that there are few preferential or non-preferential codons existing at this position. In contrast, when CUB value is much more deviation than $RSCU_0$, codons with CUB are preferentially chosen at a particular position.

Correlation analysis: In order to evaluate the context-dependent codon usage bias for the cleavage sites in FMDV polyprotein, researchers analyzed the relationship between Codon Bias Model for each position in the target regions flanking the cleavage site and the model for the corresponding cleavage site. The correlation analysis was carried out by using the multianalysis software SPSS Version 11.5 for Windows.

RESULTS AND DISCUSSION

Calculation of relative synonymous codon usage: The optimal codons among all ORFs were all G-ended codons or C-ended ones and the latter was more popular. However, the minor codons were all A-ended, U-ended, G-ended or C-ended ones. Among minor codons, majority of the codons were U-ended A-ended codons (Table 1). The date also confirmed previous observations that GC-rich synonymous codons were more frequently chosen in GC-rich organisms, in contrast, AT-rich synonymous codons were more appeared in AT-rich organisms (Andersson and Sharp, 1996; Pan *et al*., 1998; Sau *et al*., 2006). This indicated that synonymous codon usage pattern was not a random synonymous codon variation under the composition pressure. But for amino acids Ala and Gly, the optimal and minor codons were GCC and GCG, GGC and GGG, respectively (Table 2). Because the two amino acids simultaneously contained optimal and minor codons which had strong binding energy to corresponding tRNAs. It likely implied that the synonymous codon usage pattern was shaped by a complex mechanism, rather than the more abundant charged tRNAs.

Location of the cleavage sites in FMDV ORF: In comparison with the context surrounding the cleavage sites, the primary L^{pro}/VP4 cleavage occurred at K/G pair or R/G pair; VP4/VP2 cleavage occurred at a conserved A/D pair; VP2/VP3 cleavage occurred at E/G pair or Q/G pair; VP3/VP1 cleavage occurred at Q/T pair or E/T pair; VP1/2A cleavage occurred at a wide range of pairs: Q/L pair, Q/V pair, Q/M pair, Q/T pair; the primary 2A/2B cleavage occurred at a conserved G/P pair; 2B/2C cleavage occurred at a conserved Q/L pair; 2C/3A cleavage occurred at Q/I pair or L/I pair; 3A/3B cleavage occurred at E/G pair, E/Y pair; the primary 3B/$3C^{pro}$ cleavage occurred at a wide range of pairs: E/S, E/G, E/T; $3C^{pro}$/3D cleavage occurred at conserved E/G pair. Although, the two amino acid residues which was comprised of the cleavage site were not conserved to some degree, one residue always was a conserved motif for all cleavage sites.

Table 2: Synonymous codon usage and codon usage bias in FMDV

AA[a]	Codon	RSCU	AA[a]	Codon	RSCU[b]
Ala	GCA	1.000	Leu	CUA	0.192
	GCC	1.478		**CUC**	1.945
	GCG	0.614		CUG	1.617
	GCU	0.908		CUU	1.118
Arg	AGA	1.519		UUA	0.058
	AGG	0.837		UUG	1.071
	CGA	0.321	Lys	AAA	0.843
	CGC	1.669		**AAG**	1.157
	CGG	0.743	Phe	**UUC**	1.209
	CGU	0.909		UUU	0.791
Asn	**AAC**	1.727	Pro	CCA	0.800
	AAU	0.273		**CCC**	1.266
Asp	**GAC**	1.538		CCG	0.854
	GAU	0.462		CCU	1.080
Cys	**UGC**	1.106	Ser	AGC	1.105
	UGU	0.894		AGU	0.682
Gln	CAA	0.820		UCA	0.991
	CAG	1.180		**UCC**	1.574
Glu	GAA	0.690		UCG	0.906
	GAG	1.310		UCU	0.742
Gly	GGA	1.003	Thr	ACA	0.870
	GGC	1.200		**ACC**	1.611
	GGG	0.839		ACG	0.608
	GGU	0.958		ACU	0.911
His	**CAC**	1.795	Tyr	**UAC**	1.730
	CAU	0.205		UAU	0.270
Ile	AUA	0.201	Val	GUA	0.270
	AUC	1.725		GUC	1.153
	AUU	1.074		**GUG**	1.661
				GUU	0.917

[a]AA is the Amino Acid. [b]RSCU value is a mean value of each codon for a particular amino aicd. The optimal codons for each amino acid are described in bold. The minor codons for each amino acid are remarked with underline

The similarity of Codon Bias Model for the regions flanking the cleavage site: The bars of all positions in the target context reflected the CUB values and the two positions (N = 10 and N = 11) represented the cleavage site (Fig. 1-11). Although, the CUB values of the corresponding position were variable in all examined contexts, the similarity of the Codon Bias Model for some

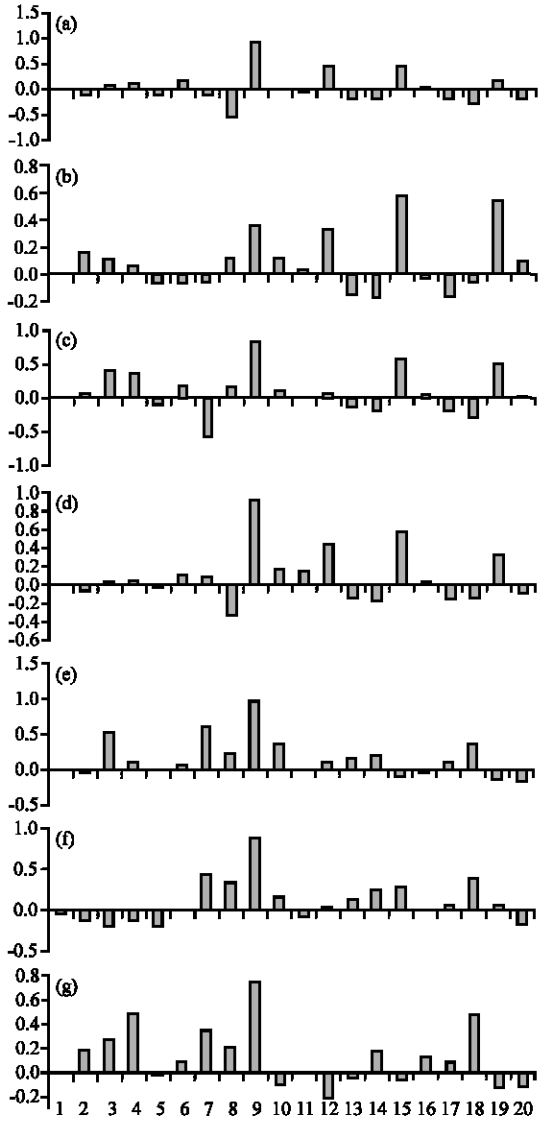

Fig. 1: The Codon Bias Model for each position in the context surrounding the cleavage site (L^{pro}/VP4) (a: Asia 1 L-VP4; b: A L-VP4; c: C L-VP4; d: O L-VP4; e: SAT 1 L-VP4; f: SAT 2 L-VP4; g: SAT 3 L-VP4)

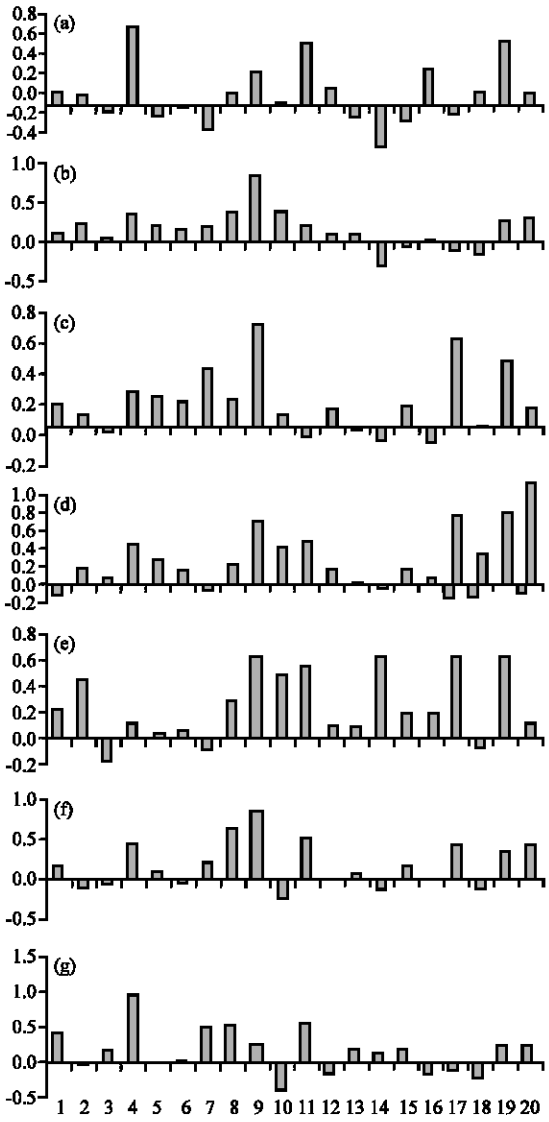

Fig. 2: The Codon Bias Model for each position in the context surrounding the cleavage site (VP4/VP2) (a: Asia 1 VP4-VP2; b: A VP4-VP2; c: C VP4-VP2; d: O VP4-VP2; e: SAT 1 VP4-VP2; f: SAT 2 VP4-VP2; g: SAT 3 VP4-VP2)

contexts were ensured to some degree (Fig. 1-11). In detail as for the L^{pro}/VP4 site, the similarity of Codon Bias Model for the corresponding position (N = 9) was observed among all serotypes (Fig. 1) as for the VP4/VP2 site, the Codon Bias Model (N = 8, 9, 19, 20) was similar, respectively (Fig. 2); for the VP2/VP3 site, the Codon Bias Model (N = 1, 2 and 19) was similar respectively (Fig. 3); for the VP1/2A site, the Codon Bias Model (N = 5, 13, 15 and 16) was similar, respectively (Fig. 5); for the 2A/2B site, Codon Bias model (N = 6, 7, 8, 9, 10, 11, 12, 13, 14, 16 and 20) was similar, respectively (Fig. 6); for the 2B/2C site, the Codon Bias Model (N = 4, 5, 10, 11 and 15) was similar, respectively (Fig. 7); for the 2C/3A site, the Codon Bias Model (N = 2, 5, 6, 9, 10, 11, 14, 15, 16 19 and 20) was similar, respectively (Fig. 8); for the 3A/3B site, the Codon Bias Model (N = 1, 8, 9, 10, 12, 13 and 18) was similar respectively (Fig. 9); for the 3B/3C site, the Codon Bias Model (N = 1, 2, 8, 9, 15, 16, 17, 18, 19 and 20) was similar respectively (Fig. 10); for the 3C/3D site, the Codon Bias Model (N = 2, 3, 4, 5, 8, 10, 11, 15, 16, 17, 18) was similar, respectively (Fig. 11), however, for the VP3/VP1 site, no

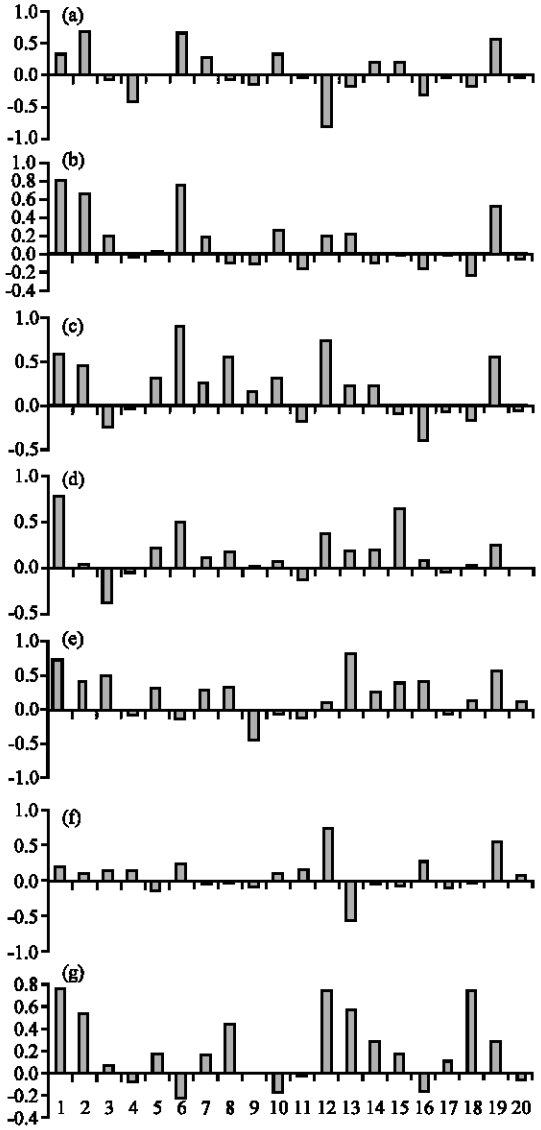

Fig. 3: The Codon Bias Model for each position in the context surrounding the cleavage site (VP2/VP3) (a: Asia 1 VP2-VP3; b: A VP2-VP3; c: C VP2-VP3; d: O VP2-VP3; e: SAT 1 VP2-VP3; f: SAT 2 VP2-VP3; g: SAT 3 VP2-VP3)

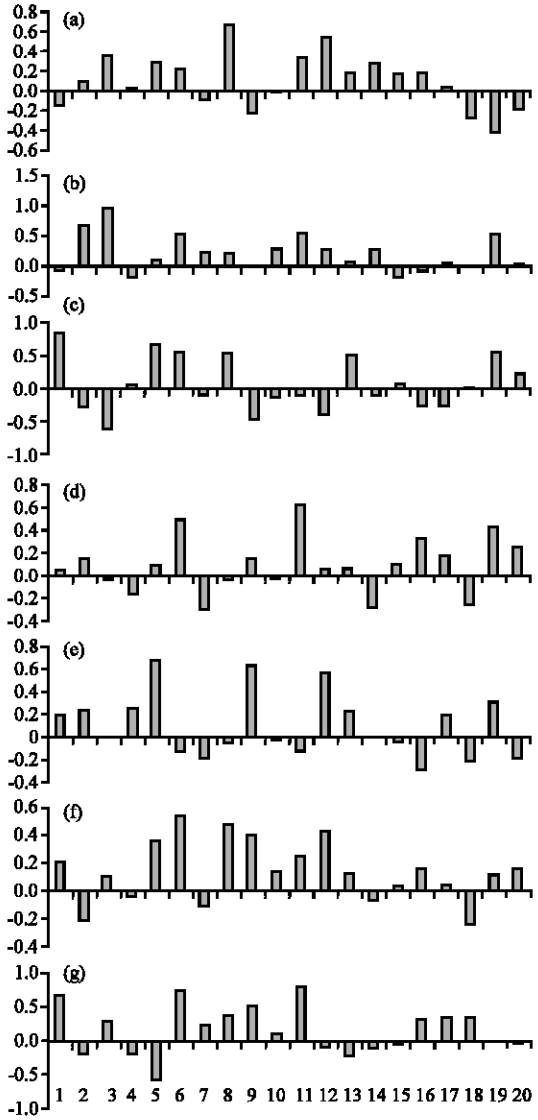

Fig. 4: The Codon Bias Model for each position in the context surrounding the cleavage site (VP3/VP1) (a: Asia 1 VP3-VP1; b: A VP3-VP1; c: C VP3-VP1; d: O VP3-VP1; e: SAT 1 VP3-VP1; f: SAT 2 VP3-VP1; g: SAT 3 VP3-VP1)

similarity of Codon Bias Model for the target region was observed (Fig. 4). Compared with the scale of the similar Codon Bias Models for the regions surrounding the L^{pro}/VP4, VP4/VP2, VP2/VP3 and VP3/VP1 sties, the scale of the similar Codon Bias Model for the regions flanking the rest cleavage sites was large. The result suggested that the degree of invariant amino acids play a role in the similarity of Codon Bias Model to some extent. It was noticed that the similarity of Codon Bias Model for the corresponding position surrounding the cleavage site likely was shaped by the Codon Bias Model for the two amino acid residues which consisted of the cleavage site.

The context-dependent codon usage bias for the cleavage site: The relationship between the Codon Bias Model for each position and the corresponding postion surrounding the cleavage site was analyzed and r value was shown in Table 1. Although, correlation between the Codon Bias Model for position in the target region flanking the cleavage site and the corresponding cleavage site among

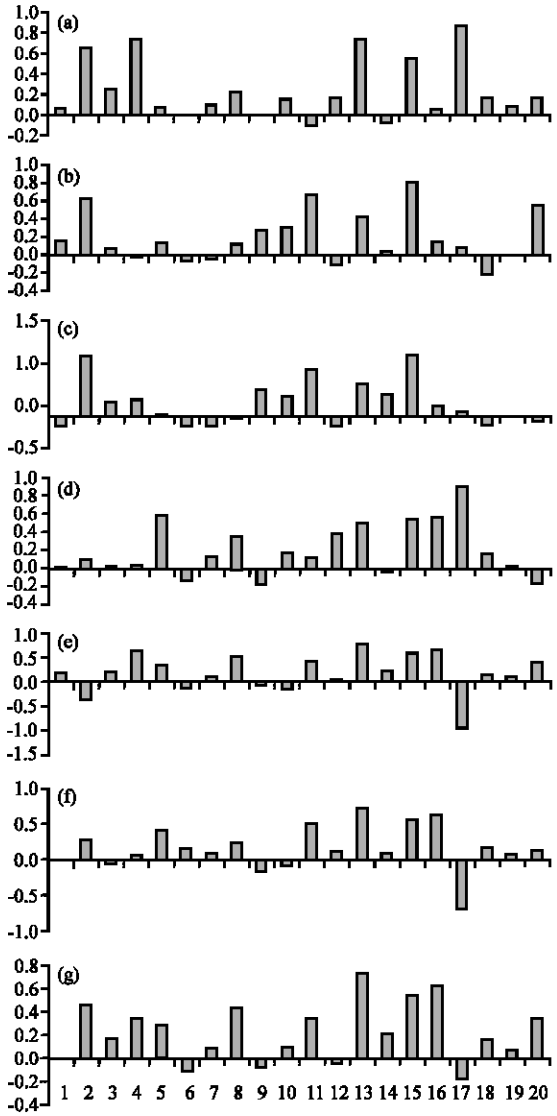

Fig. 5: The Codon Bias Model for each position in the context surrounding the cleavage site (VP1/2A) (a: Asia 1 VP1-2A; b: A VP1-2A; c: C VP1-2A; d: O VP1-2A; e: SAT 1 VP1-2A; f: SAT 2 VP1-2A; g: SAT 3 VP1-2A)

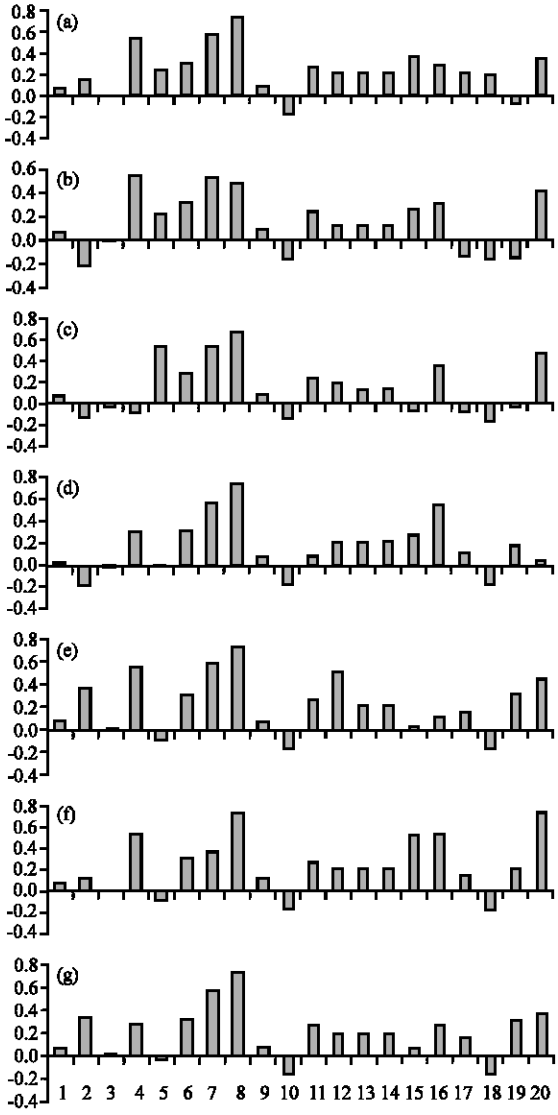

Fig. 6: The Codon Bias Model for each position in the context surrounding the cleavage site (2A/2B) (a: Asia 1 2A-2B; b: A 2A-2B; c: C 2A-2B; d: O 2A-2B; e: SAT 1 2A-2B; f: SAT 2 2A-2B; g: SAT 3 2A-2B)

all serotypes of FMDV was observed in present study, the phenomena of the context-dependent codon usage bias for the VP4/VP2, 2A/2B, 2B/2C and 3A/3B sites were ensured (Table 1).

In the study, the examined cleavage sites in the polyprotein is ensured and indicated that although these cleavage sites possess different range of amino acid pairs, one residue is always conserved. These observations may reinforce the previous views that FMDV $3C^{pro}$ can cleave all cleavage sites excluding the L^{pro}/VP4 and 2A/2B because $3C^{pro}$ needs a conserved motif served as a signal to cleave the cleavage site accurately. It is accepted that the genome of bacterial carries larger information to regulate the transcription and translation of the corresponding gene than that of the most viruses thus whether the function of influencing the translation/ translation efficiency is compressed in the brief genome of virus by the Codon Bias Model. For many reports on FMDV there is little information about the Codon Bias Model for the context surrounding the cleavage sites. There is general agreement that although 58% of the amino acids encoded by FMDV ORF are invariant but

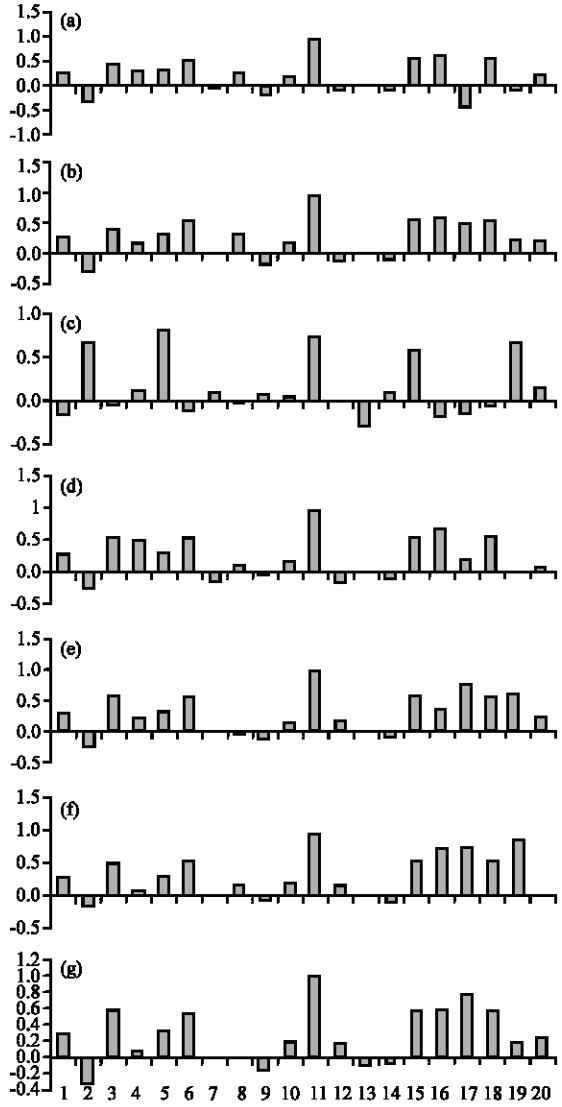

Fig. 7: The Codon Bias Model for each position in the context surrounding the cleavage site (2B/2C) (a: Asia 1 2B-2C; b: A 2B-2C; c: C 2B-2C; d: O 2B-2C; e: SAT 1 2B-2C; f: SAT 2 2B-2C; g: SAT 3 2B-2C)

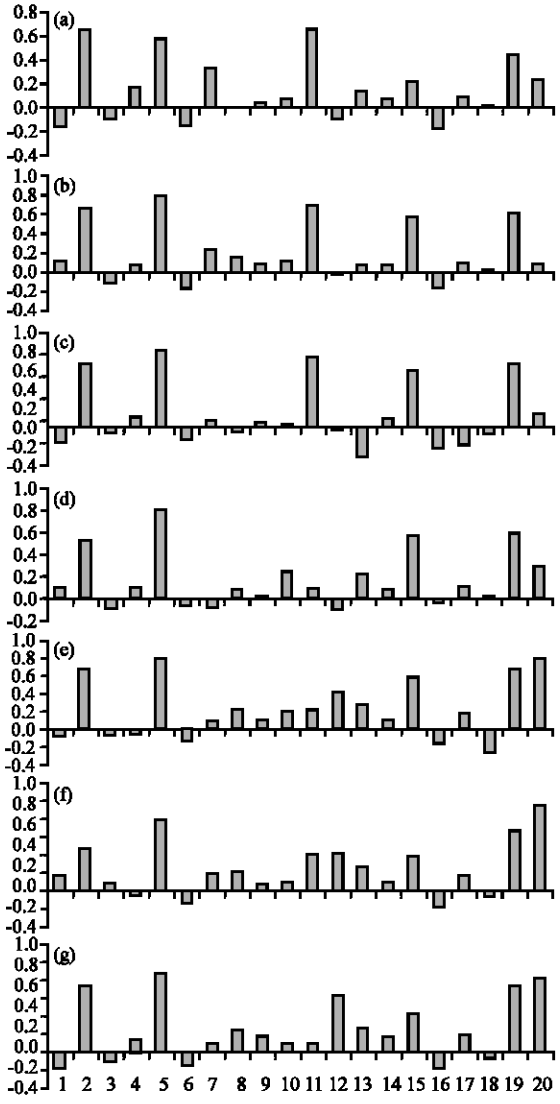

Fig. 8: The Codon Bias Model for each position in the context surrounding the cleavage site (2C/3A) (a: Asia 1 2C-3A; b: A 2C-3A; c: C 2C-3A; d: O 2C-3A; e: SAT 1 2C-3A; f: SAT 2 2C-3A; g: SAT 3 2C-3A)

variety of synonymous or non-synonymous codon usage is observed (Carrillo *et al*., 2005). Although, there are the conserved residues surrounding the cleavage sites to some degree, researchers find that the similarity of Codon Bias Model for the target context is formed without associated with the conserved residues surrounding the cleavage site, suggesting that the Codon Bias Model likely affect the translation efficiency while the Codon Bias Model for residues close to the cleavage site may influence the cleavage process.

It is known that synonymous codon usage causes different translational rates and negative CUB for codons (rarely used codons) at some critical sequences can affect translation efficiency to regulate gene expression (Zhou *et al*., 2011, 2010a, b; Chen and Inouye, 1990; Varenne *et al*., 1984). In this study, researchers find that the Codon Bias Model for the most, not all, cleavage sites which are processed by L^{pro} or $3C^{pro}$ does not have correlation with the model for its flanking regions. This result likely suggests that the contexts surrounding these cleavage sites are not associated with influencing the cleavage process. It is interesting that the VP4/VP2, 2A/2B, 2B/2C and 3A/3B sites have a common

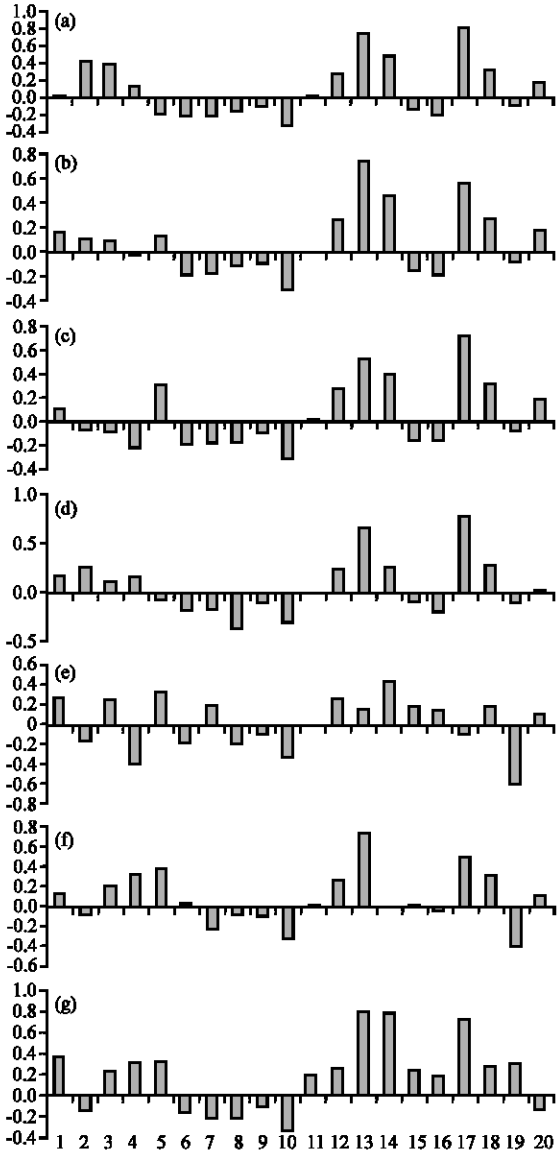

Fig. 9: The Codon Bias Model for each position in the context surrounding the cleavage site (3A/3B) (a: Asia 1 3A-3B; b: A 3A-3B; c: C 3A-3B; d: O 3A-3B; e: SAT 1 3A-3B; f: SAT 2 3A-3B; g: SAT 3 3A-3B)

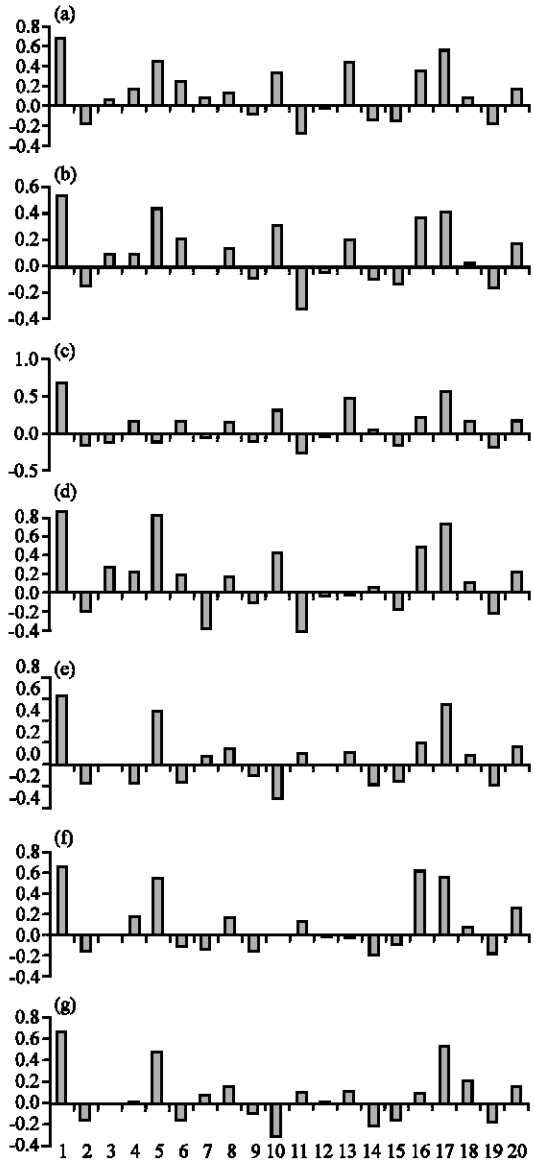

Fig. 10: The Codon Bias Model for each position in the context surrounding the cleavage site (3B/3C^{pro}) (a: Asia 1 3B-3C; b: A 3B-3C; c: C 3B-3C; d: O 3B-3C; e: SAT 1 3B-3C; f: SAT 2 3B-3C; g: SAT 3 3B-3C)

characteristic of cooperative effects by Codon Bias Model corresponding to the sites and its flanking regions on the cleavage sites. For the relationship between context-dependent codon bias and the VP4/VP2 site there are Codon Bias Model for six continuous positions downstream of the cleavage site associating with the model for the site. The process of cleavage of the VP4/VP2 site plays a key role in the transition of the provirion to the mature viron but the mechanism is completely unknown, except self-cleavage (Arnold *et al*., 1987; Curry *et al*., 1995, 1997; Harber *et al*., 1991; Lee *et al*., 1993). Some maturation of the VP4/VP2 site did not enhance the degree of intermediate stability to some degree. The Codon Bias Model for the serial positions probably contributes on this intermediate in possessing non-stability. For the relationship between context-dependent codon bias and that of the 2A/2B site there is the Codon Bias Model for four continuous positions upstream flanking this site associating with the model for

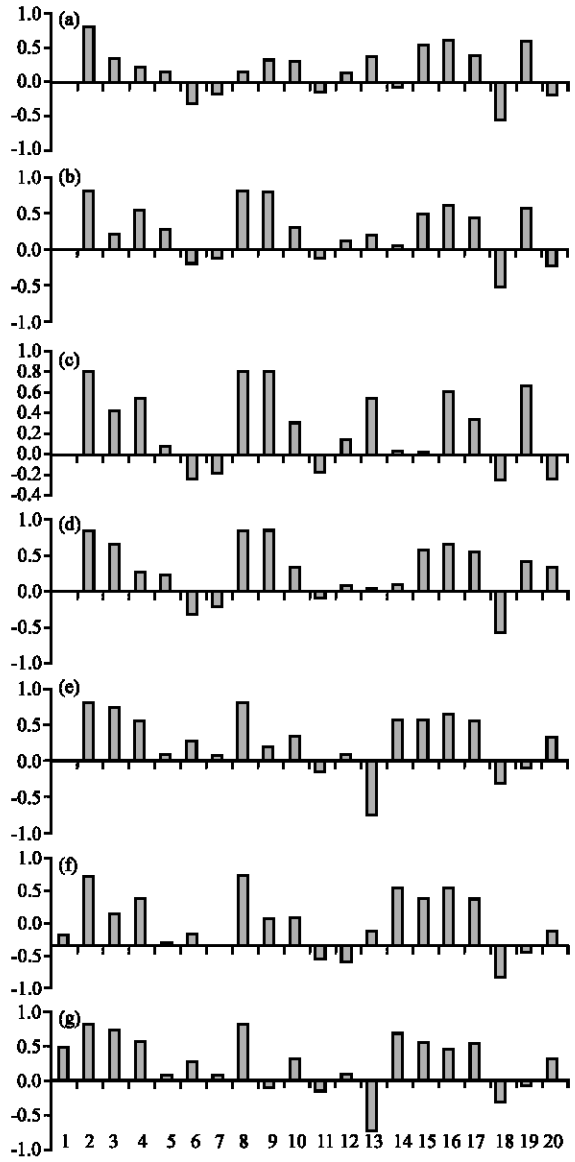

Fig. 11: The Codon Bias Model for each position in the context surrounding the cleavage site ($3C^{pro}$/3D) (a: Asia 1 3C-3D; b: A 3C-3D; c: C 3C-3D; d: O 3C-3D; e: SAT 1 3C-3D; f: SAT 2 3C-3D; g: SAT 3 3C-3D)

the site. It is accepted that proteolytic event does not take place at the 2A/2B site, the cleavage event is performed based on the translational machinery by the 2A peptide which allows the release of the protein-2A from the ribosome while permitting the synthesis of the downstream proteins to proceed (Donnelly *et al*., 2001). It is no doubt that the conserved pair (G/P) for the 2A/2B site plays partially a role in this cleavage site due to many G/P pairs existing in FMDV. This implies that the context surrounding this specific G/P residue probably mediates the cleavage case. In this study, the context-dependent codon bias for the region downstream flanking the 2A/2B site is not observed while the Codon Bias Model for four positions upstream of this site has a significant correlation with the conserved residues. For the relationship between context-dependent codon bias and the 2B/2C, 3A/3B sites, the characteristic probably imply that the Codon Bias Model mediates the translation efficiency to regulate the polypeptide elongation. For FMDV its product (L^{pro}) can ruin eIF4G arising from shut-off of host cap-dependent mRNA translation for host cell. In this situation, the amount of all kinds of tRNA was possibly excessive to translate FMDV proteins. So the Codon Bias Model for the specific region in FMDV ORF likely mediates gene transcription or expression to some degree. It is generally accepted opinion that codon bias model could regulate the expression levels of individual genes by modulating the rates of polypeptide elongation (Zhou *et al*., 2010b; Miyasaka, 1999; Ohno *et al*., 2001; Bonekamp *et al*., 1985; Chavancy and Garel, 1981; Gouy and Gautier, 1982; Kudla *et al*., 2009; Robinson *et al*., 1984; Sorensen *et al*., 1989) although, the Codon Bias Model influencing the translation efficiency for the initiation process is mainly analyzed. Sorensen *et al*. (1989) reported that optimal codons with positive CUB could enhance gene expression efficiency compared to the rarely used codons with negative CUB (Sorensen *et al*., 1989). It is noticed that due to the limit for virus genome size, virus probably compresses rich genetic information into the limited genome during evolution. The translational regulation is not only controlled under the 5' and 3' non-translational region but also influenced by some regions or sites in the coding sequence probably.

CONCLUSION

Given the results of the present study and the earlier reported experimental findings as well, the occurrence of context-dependent codon bias for the VP2/VP4 and 2A/2B sites obviously has a dramatic positive effect on the cleavage process. The reason for the context-dependent codon bias for 2B/2C and 3A/3B and few reported experimental finding is not clear; more detailed experimental approaches are required in the future. However, there is a significant correlation between the Codon Bias Model and a series of positions surrounding these cleavage sites suggesting that Codon Bias Model might cumulatively affect the translation probably causing a selectively detectable change in fitness in the process of evolution although, the contribution of each position might be slight.

ACKNOWLEDGEMENTS

This research was supported in part by grants from the Key Technologies R&D Program of Gansu Province and China (No.: 0801NKDA034 and no.2008BADB4B05). This study also was supported by the National Natural Science Foundation of China (No.: 30671563 and No.: 30700597).

REFERENCES

Alff-Steinberger, C. and R. Epstein, 1994. Codon preference in the terminal region of *E. coli* genes and evolution of stop codon usage. J. Theor. Biol., 168: 461-463.

Andersson, G.E. and P.M. Sharp, 1996. Codon usage in the Mycobacterium tuberculosis complex. Microbiol., 142: 915-925.

Arnold, E., M. Luo, G. Vriend, M.G. Rossmann and A.C. Palmenberg, 1987. Implications of the picornavirus capsid structure for polyprotein processing. Proc. Natl. Acad. Sci. USA., 84: 21-25.

Bablanian, G.M. and M.J. Grubman, 1993. Characterization of the foot-and-mouth disease virus 3C protease expressed in Escherichia coli. Virology, 197: 320-327.

Belsham, G.J., 1992. Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes *in vivo*. EMBO J., 11: 1105-1110.

Bonekamp, F., H.D. Andersen, T. Christensen and K.F. Jensen, 1985. Codon-defined ribosomal pausing in *Escherichia coli* detected by using the pyrE attenuator to probe the coupling between transcription and translation. Nucleic. Acids Res., 13: 4113-4123.

Brown, C.M., M.E. Dalphin, P.A. Stockwell and W.P. Tate, 1993. The translational termination signal database. Nucleic Acids Res., 21: 3119-3123.

Brown, C.M., P.A. Stockwell, M.E. Dalphin and W.P. Tate, 1994. The translational termination signal database (TransTerm) now also includes initiation contexts. Nucleic Acids Res., 22: 3620-3624.

Cao, X., I.E. Bergmann, R. Fullkrug and E. Beck, 1995. Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. J. Virol., 69: 560-563.

Carrillo, C., E.R. Tulman, G. Delhon, Z. Lu and A. Carreno *et al*., 2005. Comparative genomics of foot-and-mouth disease virus. J. Virol., 79: 6487-6504.

Chavancy, G. and J.P. Garel, 1981. Does quantitative tRNA adaptation to codon content in mRNA optimize the ribosomal translation efficiency? Proposal for a translation system model. Biochimie, 63: 187-195.

Chen, G.F. and M. Inouye, 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. Nucleic Acids Res., 18: 1465-1473.

Clarke, B.E. and D.V. Sangar, 1988. Processing and assembly of foot-and-mouth disease virus proteins using subgenomic RNA. J. Gen. Virol., 69: 2313-2325.

Cooke, J.N. and K.M. Westover, 2008. Serotype-specific differences in antigenic regions of foot-and-mouth disease virus (FMDV): A comprehensive statistical analysis. Infect Genet. Evol., 8: 855-863.

Curry, S., C.C. Abrams, E. Fry, J.C. Crowther, G.J. Belsham, D.I. Stuart and A.M. King, 1995. Viral RNA modulates the acid sensitivity of foot-and-mouth disease virus capsids. J. Virol., 69: 430-438.

Curry, S., E. Fry, W. Blakemore, R. Abu-Ghazaleh and T. Jackson *et al*., 1997. Dissecting the roles of VP0 cleavage and RNA packaging in picornavirus capsid stabilization: The structure of empty capsids of foot-and-mouth disease virus. J. Virol., 71: 9743-9752.

Donnelly, M.L., D. Gani, M. Flint, S. Monaghan and M.D. Ryan, 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. J. Gen. Virol., 78: 13-21.

Donnelly, M.L., G. Luke, A. Mehrotra, X. Li, L.E. Hughes, D. Gani and M.D. Ryan, 2011. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction but a novel translational effect: A putative ribosomal 'skip'. J. Gen. Virol., 82: 1013-1025.

Donnelly, M.L., L.E. Hughes, G. Luke, H. Mendoza, E. ten Dam, D. Gani and M.D. Ryan, 2001. The cleavage activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring 2A-like sequences. J. Gen. Virol., 82: 1027-1041.

Gouy, M. and C. Gautier, 1982. Codon usage in bacteria: Correlation with gene expressivity. Nucleic Acids Res., 10: 7055-7074.

Grubman, M.J. and B. Baxt, 2004. Foot-and-mouth disease. Clin. Microbiol. Rev., 17: 465-493.

Grubman, M.J., M.P. Moraes, F. Diaz-San Segundo, L. Pena and T. de los Santos, 2008. Evading the host immune response: How foot-and-mouth disease virus has become an effective pathogen. FEMS Immunol Med. Microbiol., 53: 8-17.

Harber, J.J., J. Bradley, C.W. Anderson and E. Wimmer, 1991. Catalysis of poliovirus VP0 maturation cleavage is not mediated by serine 10 of VP2. J. Virol., 65: 326-334.

Hooper, S.D. and O.G. Berg, 2000. Gradients in nucleotide and codon usage along Escherichia coli genes. Nucleic Acids Res., 28: 3517-3523.

Jackson, A.L., H. O'Neill, F. Maree, B. Blignaut, C. Carrillo, L. Rodriguez and D.T. Haydon, 2007. Mosaic structure of foot-and-mouth disease virus genomes. J. Gen. Virol., 88: 487-492.

Joern, K., 2009. Understanding the molecular epidemiology of foot-and-mouth-disease virus. Infect. Genet. Evol., 9: 153-161.

Kudla, G., A.W. Murray, D. Tollervey and J.B. Plotkin, 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. Sci., 324: 255-258.

Lee, W.M., S.S. Monroe and R.R. Rueckert, 1993. Role of maturation cleavage in infectivity of picornaviruses: Activation of an infectosome. J. Virol., 67: 2110-2122.

Lewis-Rogers, N., D.A. McClellan and K.A. Crandall, 2008. The evolution of foot-and-mouth disease virus: Impacts of recombination and selection. Infect. Genet. Evol., 8: 786-798.

Lloyd, A.T. and P.M. Sharp, 1992. Evolution of codon usage patterns: The extent and nature of divergence between Candida albicans and Saccharomyces cerevisiae. Nucleic Acids Res., 20: 5289-5295.

Lopez de Quinto, S., M. Saiz, D. de la Morena, F. Sobrino and E. Martinez-Salas, 2002. IRES-driven translation is stimulated separately by the FMDV 3'-NCR and poly(A) sequences. Nucleic Acids Res., 30: 4398-4405.

Medina, M., E. Domingo, J.K. Brangwyn and G.J. Belsham, 1993. The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. Virology, 194: 355-359.

Miyasaka, H., 1999. The positive relationship between codon usage bias and translation initiation AUG context in *Saccharomyces cerevisiae*. Yeast, 15: 633-637.

Ohno, H., H. Sakai, T. Washio and M. Tomita, 2001. Preferential usage of some minor codons in bacteria. Gene, 276: 107-115.

Pan, A., C. Dutta and J. Das, 1998. Codon usage in highly expressed genes of Haemophillus influenzae and Mycobacterium tuberculosis: translational selection versus mutational bias. Genetics, 215: 405-413.

Pariente, N., A. Airaksinen and E. Domingo, 2003. Mutagenesis versus inhibition in the efficiency of extinction of foot-and-mouth disease virus. J. Virol., 77: 7131-7138.

Robinson, M., R. Lilley, S. Little, J.S. Emtage and G. Yarranton *et al*., 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res., 12: 6663-6671.

Ryan, M.D. and J. Drew, 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artifical polyprotein. EMBO J., 13: 928-933.

Ryan, M.D., A.M. King and G.P. Thomas, 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J. Gen. Virol., 72: 2727-2732.

Sau, K., S.K. Gupta, S. Sau, S.C. Mandal and T.C. Ghosh, 2006. Factors influencing synonymous codon and amino acid usage biases in Mimivirus. Biosystems, 85: 107-113.

Sharp, P.M. and M. Bulmer, 1988. Selective differences among translation termination codons. Genet., 63: 141-145.

Sharp, P.M., T.M.F. Tuohy and K.R. Mosurski, 1986. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res., 14: 5125-5143.

Sorensen, M.A., C.G. Kurland and S. Pedersen, 1989. Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol., 207: 365-377.

Vakharia, V.N., M.A. Devaney, D.M. Moore, J.J. Dunn and M.J. Crubman, 1987. Proteolytic processing of foot-and-mouth disease virus polyproteins expressed in a cell-free system from clone-derived transcripts. J. Virol., 61: 3199-3207.

Varenne, S., J. Buc, R. Lloubes and C. Lazdunski, 1984. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Mol. Biol., 180: 549-576.

Zhou, J., J. Zhang, Y. Ding, H. Chen, L. Ma and Y. Liu, 2010a. Characteristics of codon usage bias in two regions downstream of the initiation codons of foot-and-mouth disease virus. Biosystems, 101: 20-28.

Zhou, J.H., J. Zhang, H.T. Chen, L.N. Ma and Y.S. Liu, 2010b. Analysis of synonymous codon usage in foot-and-mouth disease virus. Vet. Res. Commun., 34: 393-404.

Zhou, J.H., J. Zhang, H.T. Chen, L.N. Ma, Y.Z. Ding, Z. Pejsak and Y.S. Liu, 2011. The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus. Infect Genet Evol., 11: 1815-1819.