Journal of Animal and Veterinary Advances 11 (19): 3627-3634, 2012
ISSN: 1680-5593

Characteristics of Nucleotide and Codon Usage Bias of 2A Sequence of Foot and Mouth Disease Virus

[1]Zong-Liang Gao, [2]Yuan-Xing Gu, [1]Jian-Hua Zhou, [1]Yao-Zhong Ding,
[1]Jie Zhang, [1]Hao-Tai Chen, [1]Li-Na Ma and [1]Yong-Sheng Liu
[1]State Key Laboratory of Veterinary Etiological Biology, Chinese Academy of Agricultural Sciences,
Lanzhou Veterinary Research Institute, 730046 Lanzhou, Gansu, P.R. China
[2]Key Laboratory of Preventive Veterinary Medicine, Institute of Preventive Veterinary Medicine,
Zhejiang University, Hangzhou, Zhejiang Province, P.R. China

Abstract: The cleavage at 2A/2B junction of Foot and Mouth Disease Virus (FMDV) is mediated by the 2A sequence. Here researchers calculate some data of the 2A sequence including the nucleotide usage bias (R-value), the Authentic Codon Bias (ACB value) and accumulation of Codon Bias (CUB_{2A} value) to estimate potential effects of nucleotide and codon bias on the auto-cleavage of this junction. Researchers find that a high nucleotide usage bias exists in the 2A sequence (R-value = 0.856±0.113), ten positions 3, 5, 8, 9, 11, 14-18 show a stable codon usage tendency in the 2A sequence across all serotype while the accumulation of Codon Bias (CUB_{2A} = 0.293±0.043) is low. It is interesting that an obvious downward trend of accumulation of codon bias in C-termini is present in the 2A sequence. These phenomena suggest that the nucleotide and codon usage models play potential roles in impairing the formation of the peptide bond linking the 2A/2B junction during co-translation by affecting conformation changes of A and P sites in the ribosome.

Key words: Foot and mouth disease virus, 2A/2B junction, accumulation of codon bias, conformation change, co-translation

INTRODUCTION

Foot and mouth disease is generally considered to be the most contagious of all the diseases infecting farm animals such as cattle, pigs, sheep and goats. Foot and Mouth Disease Virus (FMDV) is classified as a number of the *Aphthovirus* genus of the family Picornaviridae. The virus exists in the form of seven different serotypes: O, A, C, Asia 1, SAT 1, SAT 2 and SAT 3 and many subtypes had evolved within each serotype. This virus possesses the single stranded positive-sense RNA genome which is used as an mRNA and encodes a viral polyprotein. Excluding 5 and 3 untranslated regions, FMDV genome contains a signal Open Reading Frame (ORF) encodes 12 products: the Leader protease (L^{pro}), four capsid proteins (VP4, VP2, VP1 and VP3) and the non-structural proteins (2A, 2B, 2C, 3A, 3B, 3C and 3D). Although, the 2A sequence is only 54 nucleotides in length and does not contain protease motifs breakage of the viral polyprotein at the 2A/2B junction is rather rapid without any protease reactions (Ryan and Drew, 1994; Vakharia *et al*., 1987). In addition, the 2A/2B cleavage of the cardiovirus Theiler's Murine Encephalomyelitis Virus (TMEV) and Encephalomyocarditis Virus (EMCV) are also mediated by their 2A proteins and the C-terminal regions of TMEV and EMCV are highly similar to the small 2A protein of FMDV (Donnelly *et al*., 1997). This view implied that the specific 2A region serves as a signal to initiate breakage at the 2A/2B junction and the highly conserved amino acid residues plays an important role in this case. Interestingly, the automatic cleavage only occurs in a Eukaryotic system and never occurs in a prokaryotic one (Donnelly *et al*., 1997). This phenomenon indicated that although ribosomes of both Eukaryotic and Prokaryotic Systems can scan and translate 2A sequence, the automatic cleavage can be affected by the different translation mechanisms between Eukaryotic and Prokaryotic Systems. Further studies showed that the proteins upstream of the 2A sequence were more translated than the products downstream of the 2A sequence in the artificial polyprotein (Donnelly *et al*., 2001a, b).

The 2A/2B junction may never be synthesized normally since, the synthesis of the specific peptide bond at this junction is prevented by the modification function of 2A sequence and the modification of the translational

Corresponding Author: Yong-Sheng Liu, State Key Laboratory of Veterinary Etiological Biology, Chinese Academy of Agricultural Sciences, Lanzhou Veterinary Research Institute, 730046 Lanzhou, Gansu, P.R. China

machinery by the 2A sequence allows the release of the protein-2A from the ribosome while permitting the synthesis of the downstream proteins to proceed (Donnelly *et al*., 2001a, b). Based on this function, the 2A sequence has been used in a wide range of heterologou protein expression and it is active in all eukaryotic systems analyzed thus far (Ryan and Drew, 1994; Camon *et al*., 1997; Collins *et al*., 1998; De Rose *et al*., 2000; Mattion *et al*., 1996; Precious *et al*., 1995; Schmidt and Rethwilm, 1995; Suzuki *et al*., 2000; Thomas and Maule, 2000; Varnavski *et al*., 2000). Although, several evidences at the amino acid level indicated that some specific amino acids in C-termination of the 2A sequence of FMDV can influence the auto-cleave at the 2A/2B junction, the views explaining some of the anomalies of the 2A cleavage reaction at the nucleotide and codon levels are not yet reported. A systematic analysis of some important potential factors in influencing the automatic breakage at the 2A/2B junction including the nucleotide usage bias and codon usage model of the 2A sequence is useful to ensure some genetic characteristics which impair the formation of the peptide in the 2A/2B junction of FMDV. Here researchers employed such analytic methods targeting at 66 samples of FMDV and found that a strong nucleotide usage bias and a specific codon usage model exist in the 2A sequence of FMDV.

MATERIALS AND METHODS

Sequence data: To evade introduction of the whole coding sequence with high homology into this study, 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 (Table 1). In order to locate the position of the 2A sequence in the polyprotein of FMDV, multiple sequence alignments were performed with the Clustal W (1.7) Method of DNAstar Software (7.0) for Windows based on the earlier report by Carrillo *et al*. (2005) and the 2A sequences were extracted from the whole coding sequences with the manual editing.

The calculation of the nucleotide usage bias of the 2A sequencel: To evaluate the degree of the nucleotide usage bias of the 2A sequence, a new index (R-value) was introduced into this study based on the earlier report (Miyasaka, 1999). The R-value ranges from 0-1 and a higher value indicates a stronger nucleotide usage bias.

$$R = \sqrt[54]{\prod_{i=1}^{i=54} W_{ij}}$$

Where:

- R = The value which represents the relative adaptiveness of the nucleotide in the 2A sequence of FMDV was calculated as a geometric mean of the W_{ij} value
- W_{ij} = The value reflects the degree of nucleotide usage bias in the same position of the 2A sequences
- i = Represents positions (from the position 1st-54th) belonging to the 2A sequence
- j = Represents A, U, G or C base

The detection of authentic codon bias of the 2A sequence together with the 1st codon of the 2B sequence: To investigate Authentic Codon Bias (ACB value) of each position in the 2A sequence together with the first position in the 2B sequence, researchers provided that statistically equal and random usage of all available synonymous codons was the neutral point ($RSCU_0 = 1.00$) for the development of serotype-specific codon usage (Zhou *et al*., 2010a, b). Because the Relative Synonymous Codon Usage (RSCU) is a basic method to estimate the ratio of the observed frequency of codons to the expected frequency given that all the synonymous codons for the same amino acids are used equally and the index is never affected by amino acids composition or the abundance ratio of synonymous codon usage, RSCU value was used to calculate $RSCU_{ij}$ value by the formula following the previous report (Sharp *et al*., 1986).

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

Where:

- ACB = 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 are $RSCU_0$ according to a particular position in the target context, ACB is equal to zero. It means that there are few preferential or non-preferential codons existing at this position.

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

Sequence No.	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/Viet Nam (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

In contrast when ACB value is much more deviate than $RSCU_0$, codons with ACB value are preferentially chosen at a particular position.

The detection of the accumulation of codon usage bias for the 2A sequence: A mathematical model for the 2A sequence was developed to investigate the circumstance which is encountered by the ribosome scanning. Researchers used a formula of CUB_{2A} values to estimate effects of codon usage pattern in the 2A sequence on ribosome scanning. Because the Relative Synonymous Codon Usage (RSCU) is a basic method to estimate the ratio of the observed frequency of codons to the expected frequency given that all the synonymous codons for the same amino acids are used equally and the index is never affected by amino acids usage or the abundance Ratio of Synonymous Codon Usage (RSCU) value was used to calculate the quantitative codon bias following the previous report (Sharp *et al*., 1986):

$$CUB_{2A} = \sqrt[18]{\prod_{k=1}^{k=18} \left| RSCU_{kl} - RSCU_0 \right|}$$

Where:

CUB_{2A} = The accumulation codon bias of all codon positions in the 2A sequence

k = The codon positions (from position the 1st-18th) in this sequence

l = A codon in the corresponding position

For detection of accumulation codon bias for different lengths from N-termination to C-termination of the 2A sequence, researchers applied the formula of CUB_{2A} to calculate the values for a series of in frame deletions from N-termination to C-termination, namely $CUB_{2A\,1\text{-}18}$, $CUB_{2A\,2\text{-}18}$, $CUB_{2A\,3\text{-}18}$, ..., $CUB_{2A\,16\text{-}18}$, $CUB_{2A\,17\text{-}18}$, $CUB_{2A\,18}$.

RESULTS AND DISCUSSION

The nucleotide usage bias of the 2A sequence: The data of nucleotide usage bias (R-value) of the 2A sequence is an average of 0.856 and SD of 0.113 indicating that the strong nucleotide bias exists in the 2A sequence (Table 2). It implies that although low mutation rate is present in the 2A sequence (Donnelly *et al*., 1997; Carrillo *et al*., 2005) structural and functional constraints shape the nucleotide usage model in this short region.

The authentic codon bias of each position of the 2A sequence: Although, ACB values for the corresponding codon position in the 2A sequence are variable in all serotypes of FMDV, the stability of codon bias for some

Table 2: The nucleotide usage bias and accumulation of codon bias of the 2A sequence of FMDV

No.[a]	Nucleotide usage bias (R-value)	Accumulation of Codon Bias ($CUB_{2A\,1\text{-}18}$ value)[b]
1	0.942	0.309
2	0.864	0.283
3	0.794	0.319
4	0.865	0.151
5	0.850	0.282
6	0.821	0.309
7	0.850	0.152
8	0.888	0.263
9	0.914	0.305
10	0.970	0.320
11	0.877	0.334
12	0.935	0.348
13	0.818	0.280
14	0.790	0.346
15	0.780	0.263
16	0.775	0.295
17	0.812	0.367
18	0.824	0.294
19	0.806	0.298
20	0.871	0.294
21	0.936	0.305
22	0.936	0.366
23	0.885	0.305
24	0.939	0.319
25	0.867	0.274
26	0.779	0.354
27	0.936	0.305
28	0.894	0.341
29	0.856	0.323
30	0.894	0.341
31	0.748	0.318
32	0.840	0.300
33	0.894	0.346
34	0.932	0.320
35	0.933	0.267
36	0.924	0.323
37	0.933	0.267
38	0.956	0.323
39	0.909	0.339
40	0.926	0.291
41	0.815	0.240
42	0.956	0.323
43	0.956	0.323
44	0.744	0.253
45	0.924	0.297
46	0.924	0.297
47	0.815	0.240
48	0.933	0.267
49	0.870	0.272
50	0.900	0.244
51	0.831	0.223
52	0.924	0.323
53	0.926	0.291
54	0.856	0.314
55	0.824	0.287
56	0.767	0.217
57	0.833	0.290
58	0.861	0.314
59	0.786	0.272
60	0.887	0.320
61	0.800	0.308
62	0.890	0.277
63	0.770	0.266
64	0.850	0.281
65	0.850	0.281
66	0.812	0.199

[a]Represents 66 strains of FMDV in this study; [b]Represents the accumulation of codon bias of the complete 2A sequence of FMDV

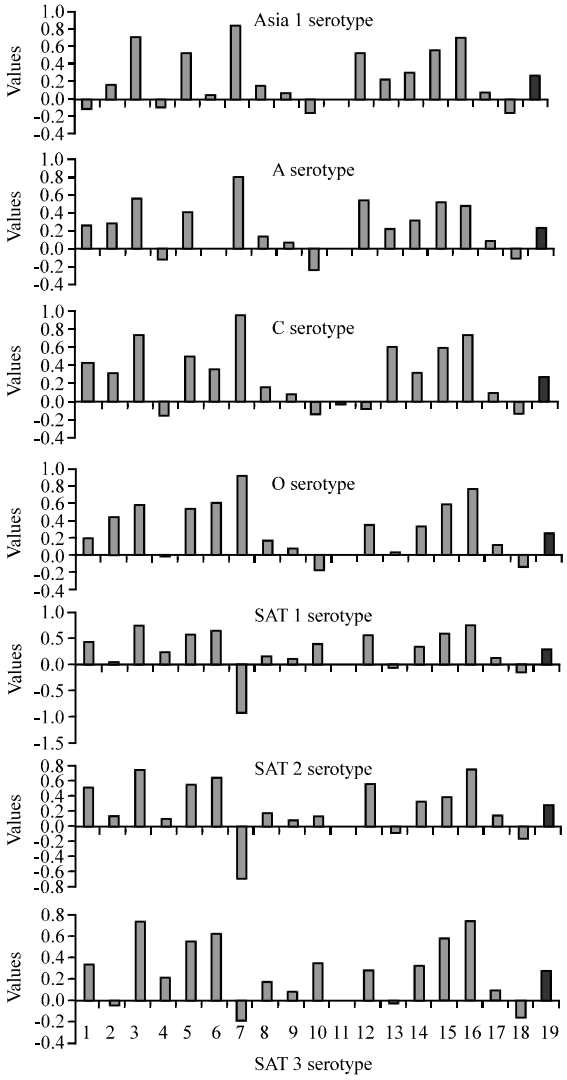

Fig. 1: The authentic codon bias for each position in the 2A sequence together with the first codon of the 2B sequence. The grey bars belong to the 2A sequence and the black bar presents the first codon of the 2B sequence

positions exists in different serotypes to some degree. In detail, the stable codon bias exists in the 3rd, 5th, 8th, 9th, 11th, 14th, 15th, 16th, 17th, 18th positions of the 2A sequence across all serotypes (Fig. 1). Compared with the conserved residues in the positions 6th, 7th, 9th-14th, 16th-18th in the 2A sequence (Carrillo *et al*., 2005), the codon positions with the stable codon bias in this region overlap the conserved residues implying that these overlapping positions have a potential effect on the biological functions of the 2A protein.

The accumulation of codon bias of the 2A sequence: The data of the accumulation of Codon Usage Bias

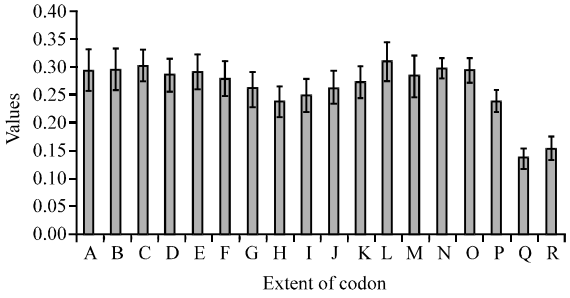

Fig. 2: The extents of codon bias accumulation of different lengths of the 2A peptide. The grey bars represent the extents of codon bias scanning from N-termini to C-termination of the 2A peptide. The A bar represents the extent of codon bias accumulation from the 1st position to 18th; the B bar represents the extent from the 2nd position to 18th; the C bar represents the extent from the 3rd position to 18th; the D bar represents the extent from the 4th-18th; the E bar represents the extent from the 5th-18th; the F bar represents the extent from the 6th-18th; the G bar represents the extent from the 7th-18th; the H bar represents the extent from the 8th-18th; the I bar represents the extent from the 9th-18th; the J bar represents the extent from the 10th-18th; the K bar represents the extent from the11th-18th; the L bar represents the extent from the 12th-18th; the M bar represents the extent from the 13th-18th; the N bar represents the extent from the 14th-18th; the O bar represents the extent from 15th-18th; the P bar represents the extent from 16th-18th; the Q bar represents the extent from 17th-18th

(CUB_{2A} value) of the complete 2A sequence is an average of 0.293 and SD of 0.043 indicating that the low codon bias exists in the 2A sequence (Table 2). The low CUB_{2A} value indicates that this peptide is composed of amino acids which are encoded by synonymous codons with weak codon bias and the pattern of codon bias of this peptide bears resemblance to the overall pattern of the whole coding sequence of FMDV. These characteristics are shaped under the limitation of the overall codon bias of FMDV. The various extents of accumulation of codon bias by a series of in-frame deletions of all residues of the 2A sequence were calculated (Fig. 2). These serial values for the accumulation of codon bias reflect that when ribosomes scan the C-termini of the 2A sequence, the downward trend of accumulation of codon bias have a potential effect on influencing the formation of the bond of the 2A/2B junction by the interaction between ribosome and this part of the 2A sequence.

For FMDV polyprotein maturation, the main activities are mediated by proteolytic processing of 3C protease. Except for proteolytic activities, the autonomous breakage of the 2A/2B junction is a representative genetic characteristic. Although, some experiments indicated that the 2A sequence plays an important role in breakage of this junction (Ryan and Drew, 1994; Donnelly *et al*., 1997), molecular mechanism accounting for some of the anomalies of the 2A cleavage reaction awaits further clarification. Although, FMDV genome has a high mutation rate (Carrillo *et al*., 2005; Zhou *et al*., 2010a, b; Zhong *et al*., 2007), the 2A sequence contains a strong nucleotide usage preference (average R-value = 0.855) under selection pressure. It may be due to that viruses are adept at manipulating the molecular events occurring during the elongation cycle of protein synthesis. Some findings also indicated that the coding sequences contain an integral message for the various molecular mechanisms involved in gene expression (Gustafsson *et al*., 2004; Komar *et al*., 1999; Sharp *et al*., 2010). At the translation level, the 2A peptide is linked with the structural protein and this region enables its upstream to translate greater molar excess than its downstream (Donnelly *et al*., 2001a), the nucleotide and codon usage patterns of the 2A sequence likely serve as a valve to regulate the translation level of the structural protein.

Degeneracy in the genetic code allows an additional potential for the coding sequence to carry structural information regarding the encoded protein that can be at the level of a single codon or at a contiguous nucleotide region (Brunak *et al*., 1994). Compared with weak codon bias (average CUB_{2A} = 0.293) of the 2A sequence caused by the general codon bias of FMDV, the strong nucleotide preference (R = 0.855) of the 2A sequence possibly serve as a regulation signal for ribosomes. In this study, the 2A sequence possesses dominant positions with positive codon bias and the authentic codon bias regularly drops from the 16th-18th position across all serotypes of FMDV; the downward trend of the accumulation of codon bias significantly descends at the C-termination of this region. These data may suggest that allosterism between A site and P site at the C-termination of this region contributes to non-formation of the specific peptide bond at this junction. Programmed ribosomal frame-shifting and programmed ribosomal hopping are also harnessed in the expression of certain cellular genes (Farabaugh, 1996). Due to an allosteric effect of the A and P sites in the ribosome, tRNAs binding to the two sites represents an arrangement of angle rather than linear and conformational changes in A-sitet RNA and P-site result from ribosome scanning the mRNA context (Green and Noller, 1997). The conformational changes could result from the context mediated by the patterns of nucleotide and codon usage. In detail, the codon usage and the nucleotide usage patterns in coding sequence of mRNA could affect the allosteric effect of the two sites since, the precise structures resulted from various base composition in coding sequences may influence the interaction between codons and anti-codons and regulate the process of delivery of tRNA from A site to P site.

Peptidyl transferase then catalyzes peptide bond formation between the nascent polypeptide chain carried by the peptidel-tRNA and the incoming aminoacyl tRNA. Weygand-Durasevic and Ibba (2010) found that the nucleotide and codon usage pattern can play important roles in the protein synthesis (Weygand-Durasevic and Ibba, 2010). For the context of FMDV 2A sequence, the high nucleotide usage bias, positions with stably ACB value and the downward trend of accumulation of codon bias in the C-termini could mediate the conformational changes of two work planes (A and P sites) in the ribosome. The specific conformational change in the 2A/2B junction likely impairs the formation of the bond which links the 2A protein and 2B protein of FMDV (Table 3).

Table 3: The degree of synonymous codon usage bias of the whole coding sequence of FMDV

Amino acids	CUB	Amino acids	CUB
GCA	0.073	CUA	0.806
GCC	0.478	CUC	0.946
GCG	0.386	CUG	0.617
GCU	0.116	CUU	0.206
AGA	0.518	UUA	0.940
AGG	0.175	UUG	0.129
CGA	0.679	AAA	0.157
CGC	0.666	AAG	0.157
CGG	0.258	UUC	0.211
CGU	0.150	UUU	0.211
AAC	0.727	CCA	0.199
AAU	0.727	CCC	0.266
GAC	0.538	CCG	0.164
GAU	0.538	CCU	0.114
UGC	0.150	AGC	0.146
UGU	0.150	AGU	0.318
CAA	0.183	UCA	0.138
CAG	0.183	UCC	0.574
GAA	0.310	UCG	0.135
GAG	0.310	UCU	0.272
GGA	0.103	ACA	0.159
GGC	0.209	ACC	0.611
GGG	0.162	ACG	0.392
GGU	0.111	ACU	0.118
CAC	0.795	UAC	0.730
CAU	0.795	UAU	0.730
AUA	0.799	GUA	0.720
AUC	0.725	GUC	0.192
AUU	0.133	GUG	0.674
-	-	GUU	0.151

CONCLUSION

Taken together, researchers find that the high nucleotide usage bias exists in the 2A sequence; the ten positions 3, 5, 8, 9, 11, 14-18 show a stable codon usage tendency in the 2A sequence across all serotype and these positions do not overlap the conserved residues in this region; the downward trend of accumulation of codon bias exists in the C-termini of the 2A sequence. These genetic characteristics could reflect effects of the nucleotide and codon usage biases on impairing the formation of the bond linking between the 2A and 2B proteins of FMDV.

ACKNOWLEDGEMENTS

This research was supported in parts by grants from National Science and Technology Key Project (2009ZX08007-006B) and International Science and Technology Cooperation Program of China (No. 2010 DFA32640) and Science and Technology Key Project of Gansu Province (No. 0801NKDA034). This study was also supported by National Natural Science foundation of China (No. 30700597 and No. 31072143). Zong-Liang Gao and Yang-Xiang Gu contributed equally in this research.

REFERENCES

Brunak, S., J. Engelbrecht and C. Kesmir, 1994. Correlation Between Protein Secondary Structure and The mRNA Nucleotide Sequence. In: Protein Structure by Distance Analysis, Bohr, H. and S. Brunak (Eds.). IOS Press, Amsterdam, pp: 327-334.

Camon, E.B., P.J. Chaplin, C.J. Howard and R.A. Collins, 1997. Modulation of bovine T cell responses by IL-12 and the influence of IL-4 on interferon-gamma respiratory syncytial virus. Biochem. Soc. Trans., 25: 291S-291S.

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.

Collins, R.A., E.B. Camon, P.J. Chaplin and C.J. Howard, 1998. Influence of IL-12 on interferon-gamma production by bovine leucocyte subsets in response to bovine respiratory syncytial virus. Vet. Immunol. Immunopathol., 63: 69-72.

De Rose, R., J.P.Y. Scheerlinck, G. Casey, P.R. Wood, J.M. Tennent and P.J. Chaplin, 2000. Ovine interleukin-12: Analysis of biologic function and species comparison. J. Interferon Cytokine Res., 20: 557-564.

Donnelly, M.L., D. Cani, M. Flint, S. Monoghan and M.D. Ryan, 1997. The cleavage activities of aphthovirus and cardiovirus 2A protein. J. Gen. Virol., 78: 1-13.

Donnelly, M.L., G. Luke, A. Mehrotra, X. Li, L.E. Hughes, D. Gani and M.D. Ryan, 2001a. Analysis of the aphovirus 2A/2B polyprotein cleavage mechanism indicates not a proteolytic reaction, but a novle translational effects: 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, 2001b. The cleavage activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring 2A-like sequences. J. Genet. Virol., 82: 1027-1041.

Farabaugh, P.J., 1996. Programmed translational frameshifting. Ann. Rev. Genet., 30: 507-528.

Green, R. and H.F. Noller, 1997. Ribosomes and translation. Ann. Rev. Biochem., 66: 679-716.

Gustafsson, C., S. Govindarajan and J. Minshull, 2004. Codon bias and heterologous protein expression. Trends Biotechnol., 22: 346-353.

Komar, A.A., T. Lesnik and C. Reiss, 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. FEBS Lett., 462: 387-391.

Mattion, N.M., E.C. Harnish, J.C. Crowley and P.A. Reilly, 1996. Foot-and-mouth disease virus 2A protease mediates cleavage in atteruated Sabin 3 poliovirus vectors engineered for delivery of foreign antigens. J. Virol., 70: 8124-8127.

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

Precious, B., D.F. Young, A. Bermingham, R. Fearns, M.P. Ryan and R.E. Ranall, 1995. Inducible expression of the p, V and NP genes of the paramyxovirus simian virus 5 in cell lines and an examination of the NP-P and NP-V interactions. J. Virol., 69: 8001-8010.

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.

Schmidt, M. and A. Rethwilm, 1995. Replication foamy virus-based vectors directing high-level expression of foreign genes. Virology, 210: 167-178.

Sharp, P.M., L.R. Emery and K. Zeng, 2010. Forces that influence the evolution of codon bias. Phil. Trans. R. Soc., 365: 1203-1212.

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.

Suzuki, N., l.M. Geletka and D.L. Nuss, 2000. Essential and dispensable virus-encoded replication elements revealed by efforts to develop hypoviruses as gene expression vectors. J. Virol., 74: 7568-7577.

Thomas, C.L. and A.J. Maule, 2000. Limitatious on the use of fused green fluorescent protein to investigate structure-function relationships for the cauliflower mosaic virus movement protein. J. Gen. Virol., 81: 1851-1855.

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.

Varnavski, A.N., P.R. Young and A.A. Knromykih, 2000. Stable high-level expression of heterologous genes *in vitro* and *in vivo* by noncytopathic DNA-based Kunjin virus replicon vectors. J. Virol., 74: 4394-4403.

Weygand-Durasevic, I. and M. Ibba, 2010. New roles for codon usage. Science, 329: 1473-1474.

Zhong, J., Y. Li, S. Zhao, S. Liu and Z. Zhang, 2007. Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes, 35: 767-776.

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.