VIROLOGY JOURNAL

Mapping codon usage of the translation initiation region in porcine reproductive and respiratory syndrome virus genome

Jun-hong Su[1†], Xiao-xia Ma[2†], Ya-li He[3], Ji-dong Li[1,4], Xu-sheng Ma[1], Yong-xi Dou[1], Xue-nong Luo[1] and Xue-peng Cai[1*]

Abstract

Background: Porcine reproductive and respitatory syndrome virus (PRRSV) is a recently emerged pathogen and severely affects swine populations worldwide. The replication of PRRSV is tightly controlled by viral gene expression and the codon usage of translation initiation region within each gene could potentially regulate the translation rate. Therefore, a better understanding of the codon usage pattern of the initiation translation region would shed light on the regulation of PRRSV gene expression.

Results: In this study, the codon usage in the translation initiation region and in the whole coding sequence was compared in PRRSV ORF1a and ORFs2-7. To investigate the potential role of codon usage in affecting the translation initiation rate, we established a codon usage model for PRRSV translation initiation region. We observed that some non-preferential codons are preferentially used in the translation initiation region in particular ORFs. Although some positions vary with codons, they intend to use codons with negative CUB. Furthermore, our model of codon usage showed that the conserved pattern of CUB is not directly consensus with the conserved sequence, but shaped under the translation selection.

Conclusions: The non-variation pattern with negative CUB in the PRRSV translation initiation region scanned by ribosomes is considered the rate-limiting step in the translation process.

Keywords: PRRSV, codon usage bias, translation initiation region, translation efficiency, translation selection

Introduction

Porcine reproductive and respiratory syndrome virus (PRRSV) infection causes serious disease in swine populations with a series of clinical consequences, such as high mortality, reproductive failure, post-weaning pneumonia and growth reduction [1,2]. Based on its serological characteristics, PRRSV has two main serotypes, which named the Northern American isolate (US) and the European isolate (EU), respectively [3-7]. PRRSV is an enveloped, single-stranded positive-sense RNA virus with a genome size of about 15.4kb and classified into the order *Nidovirales* of family *Arteriviridae* [8,9]. The PRRSV genome contains

ORF1a, encoding papain-like cysteine protease, ORF1b, encoding RNA dependent RNA polymerase, ORF2-6, encoding envelop proteins, and ORF7, encoding the nucleocapsid protein [10-13]. Despite a well-organization of the ORFs within the single RNA genome, viral proteins are in fact encoded from subgenomic RNAs that are likely generated through a discontinuous transcription mechanism [12,14]. Therefore, each subgenomic RNA could be translated at different translation rates that are regulated by codon usage bias (CUB). Because the faster a polypeptide chain is completed, the more rapid the ribosomes return to initiate and complete another polypeptide chain. The relationship between the efficiency of translation initiation and the level of gene expression has been well-established in many species [15-19]. Moreover, when the distance between the initiation codon and the non-preferential site is less than 50-60 positions (codons), the

* Correspondence: caixp@vip.163.com
† Contributed equally
[1]State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046, PR China
Full list of author information is available at the end of the article

ribosomes can be blocked at the non-preferential positions to shape a queue of ribosomes [20].

It is generally considered that the alternative synonymous codons are not used with equal frequencies among organisms, and the codon usage pattern plays a role in genes expressed at higher levels [21-30]. Jacques and Dreyfus proposed that the translation initiation site is a rate-limiting factor for gene expression [31]. Nevertheless, a regulatory relationship, which is thought to be mediated by preferential codons, between CUB and translation efficiency for individual genes is challengeable [32,33]. This suggested that a heterogonous gene is not necessarily expressed at a low level simply because its codons are infrequently translated by the host cell. There is a codon bias with respect to intragenic codon bias in the initial sequences of genes for which major proteins are strikingly different from their downstream codon bias. It is found that the translational initiation region plays an important role in regulating the translational efficiency and the pattern of synonymous codon usage varies in different regions along a coding sequence [34,35]. This indicated that the alternative synonymous codon usage might be related with gene function, protein structure and translation efficiency. In this study, we focus on the pattern of CUB in the translation initiation region of PRRSV as well as the characteristics of the synonymous codon usage at each position in the target region, since the interest in the pattern of CUB has been aroused by its potential relevance to the translational efficiency of PRRSV subgenomic RNAs. And the frequency of non-preferential codons usage in the target region is investigated in order to evaluate the role of translation selection on the formation of negative CUB pattern.

2. Materials and methods
2.1. Sequences data and the synonymous codon usage value
The 13 complete RNA sequences of PRRSV were downloaded from the National Center for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/Genbank/ and the synonymous codon usage values (SCUV) for this virus were reported previously [30]. Multiple alignment analyses were performed with the Clustal W (1.7) method of DNAStar software (7.0) for windows. The translation initiation regions (the 1st to the 50th residue) of ORF1a, ORF2, ORF3, ORF4, ORF5, ORF6 and ORF7 were used as targets for alignment analysis respectively.

2.2. The calculation of codon usage bias
To calculate CUB, it is supposed 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 [19]. CUB:

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

More simply, *CUB* is the average value of difference between $RSCU_{ij}$ and $RSCU_0$ at each position of the target region. *n* represents all codons appearing in this position. When all *RSCU* values according to a particular position in the target region 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.

2.3. Analysis of codon usage characteristic of the translation initiation region
We analyzed the codon usage characteristics of the translation initiation region depending on *R* values, where the *R* value, computed as the ration $R = (n_i/N_i)/(n/N)$, represents the relative abundance for a particular codon in the translation initiation region. n_i represents the total number of a particular codon within the 1st to ith amino acids, N_i represents the total number of corresponding amino acid in the 1st to ith amino acid ones, n is the total number of a certain codon within the whole coding sequence, and N is the total number of corresponding amino acids within the whole coding sequence. When *R* value is equal to 1.00, it means that the frequency of this codon in the target region is equal to the frequency of this codon in the whole coding sequence; when *R* value is lower than zero, it implies that the frequency of this codon in the target region is lower than that of the whole coding sequence; when *R* value is higher than zero, it suggests that the frequency of this codon is higher than that of the whole coding sequence.

2.4. Aanalysis of characteristics of positions with negative CUB in the target regions
To substantiate the characteristics of codon usage for positions with negative CUB in the target regions, we analyzed the target positions depending on the data, (i) the variations of codons and amino acids, (ii) *R* values for codons of the target positions.

3. Results
3.1. Multiple alignment analysis
The consensus amino acid sequence is based on the comparison of the strains in previous study [30]. The positions of amino acid conservation are listed in Table 1. The conservation of amino acid usage in translation region was analyzed. For ORF1a, 94% of amino acids in the target region of US serotype were invariant; 70% in the target region of EU serotype were conserved. For ORF2, 78% of amino acids were invariant in US serotype;

Table 1 The positions of invariant amino acids in the translation initiation region

ORF	Serotype	The position of amino acid conservation in the translation initiation region
ORF1a	US	The 2nd to 17th, 19th to 34th, 36th to 41st, 43rd to 50th
	EU	The 3rd, 6th to 13th, 15th to 18th, 20th to 23rd, 25th to 28th, 30th to 32nd, 34th, 35th, 39th to 41st, 44th, 46th, 48th to 50th
ORF2	US	The 2nd to 4th, 6th, 8th, 11th to 13th, 15th to 22nd, 25th to 31st, 33rd to 41st, 43rd to 44th, 46th to 49th
	EU	The 2nd to 4th, 7th, 12th to 13th, 15th, 18th, 20th, 22nd, 24th to 27th, 32nd to 37th, 40th, 41st, 43rd to 49th
ORF3	US	The 4th, 5th, 7th, 9th to 12th, 14th, 16th to 19th, 21st, 22nd, 24th to 26th, 29th, 31st, 33rd to 47th, 49th, 50th
	EU	The 2nd, 4th, 15th, 18th, 20th, 24th to 26th, 28th, 31st to 50th
ORF4	US	The 2nd, 6th to 8th, 10th to 12th, 14th, 17th to 31st, 33rd, 34th, 36th to 41st, 44th, 46th to 50th
	EU	The 3rd, 4th, 6th, 7th, 9th, 12th, 13th, 17th to 32nd, 34th, 36th to 39th, 41st, 42nd, 44th, 46th to 48th, 50th
ORF5	US	The 2nd, 6th to 8th, 10th, 12th, 14th, 15th, 18th to 23rd, 26th to 28th, 30th to 34th, 36th, 39th to 46th, 48th to 50th
	EU	The 3rd, 4th, 6th, 7th, 14th to 16th, 18th, 19th, 21st, 24th, 26th to 28th, 30th to 34th, 36th, 39th to 46th, 48th to 50th
ORF6	US	The 2nd to 9th, 11th to 15th, 17th to 50th
	EU	The 2nd, 4th, 5th, 7th, 8th, 15th to 22nd, 24th to 50th
ORF7	US	The 2nd to 10th, 12th to 14th, 16th to 45th, 47th, 50th
	EU	The 2nd, 3rd, 5th, 6th, 8th to 10th, 12th, 15th to 21st, 23rd to 28th, 30th, 31st, 33rd, 35th to 38th, 42nd to 50th

60% were invariant in EU serotype. Non-conserved amino acids scattered into the target regions of both US and EU serotypes. For ORF3, 74% of amino acids were invariant in US serotype; 60% were invariant in EU serotype, the most conserved amino acids tended to exist in the C' termination of the target regions of both US and EU serotypes. For ORF4, 76% of amino acids were invariant in US serotype; 72% were invariant in EU serotype. Non-conserved amino acids scattered in the flank of the target regions of both US and EU serotypes. For ORF5, 72% of amino acids were invariant in US serotype; 66% were invariant in EU serotype. Non-conserved amino acids scattered into the target regions of both US and EU serotypes. For ORF6, 96% of amino acids were invariant in US serotype; 82% were invariant in EU serotype, and non-conserved amino acids had a tendency to exist in the N' termination. For ORF7, 90% of amino acids were invariant in US serotype; 76% were invariant in EU serotype, and conserved amino acids scattered into the target region compared with that of US serotype. The various extents of the conserved amino acids encoded by ORFs of PRRSV suggested that these residues played an important role in virus biology.

3.2. Characteristics of codon usage bias in the target regions

The bars of all positions in the translation initiation region represented the CUB degree (Figure 1). Although different invariant degrees of the amino acids exist in the target regions between US and EU serotypes, the similar patterns of codon usage are present in the target regions of both US and EU serotypes (Table 2). For ORF1a, 58% of positions possess the similar pattern of codon usage in the target regions of both serotypes. Although the two

target regions corresponding to both the US and EU serotypes have a significant difference to the conservation in obvious amino acids, a large size of the similar patterns of codon usage exist in the target region and the most positions possessed the positive codon usage bais (Figure 1A). For ORF2, 34% of positions have the similar pattern of codon usage, and the positions in the N-terminal fragment had a tendency to choose low codon bias. It was also observed that the number of the positions with the negative codon usage bias for US serotype was more than that of EU serotype (Figure 1B). For ORF3, 62% of positions have the similar pattern of codon usage (Figure 1C). For ORF4, 72% positions contain the similar pattern of codon usage (Figure 1D). For ORF5, 40% of positions have the similar pattern of codon usage, and these positions with the similar pattern of codon usage do not appear to exist near the N' termination (Figure 1E). For ORF6, 26% of positions which contain the similar pattern of codon usage do not exist near the N' termination (Figure 1F). For ORF7, 44% of positions have the similar pattern of codon usage, and the most positions with low codon usage bias tend to exist near the N-terminal fragment (Figure 1G).

The various extents of the conserved pattern of codon usage for their positions in PRRSV ORFs suggest that CUB associated with these positions might modulate the corresponding gene expression.

3.3. The rate of codon usage frequency in the translation initiation region to that of the whole coding sequence

The R value for each codon was calculated and listed in Table 3. A higher R value indicated more preferential usage in the translation initiation site than that of the whole coding sequence. CUB_{ij} value for each codon was
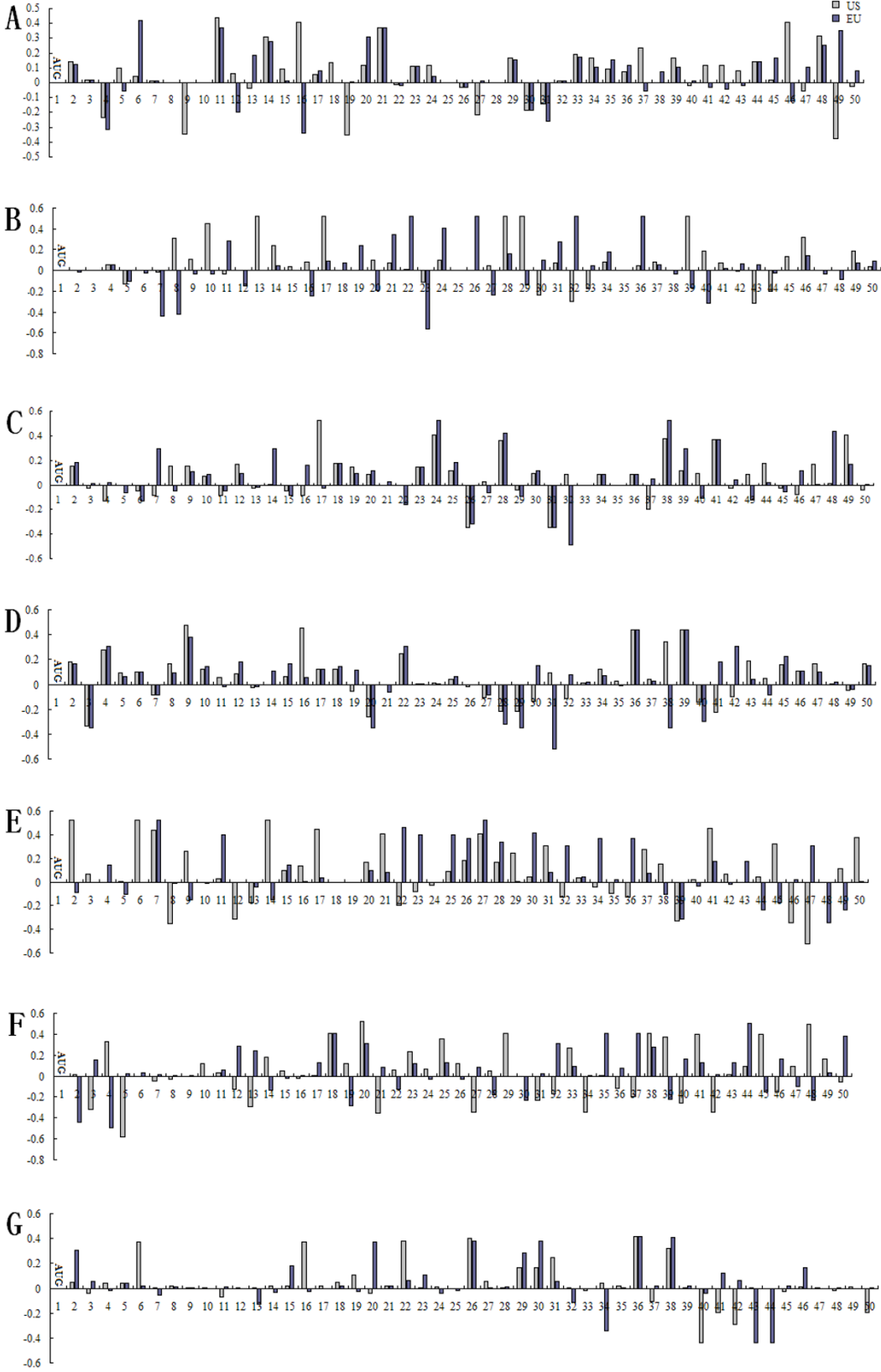
Figure 1 The CUB degree of translation initiation region in PRRSV ORFs, the white bar represents US serotype while the gray represents EU. A, ORF1a; **B**, ORF2; **C**, ORF3; **D**, ORF4; **E**, ORF5; **F**, ORF6; **G**, ORF7.

Table 2 The similar pattern of codon usage in the target regions in both US and EU serotypes

ORFs	The positions corresponding to similar codon usage pattern in the target region
ORF1a	the 2nd to 4th, 6th to 8th, 10th, 11th, 14th, 15th, 17th, 20th to 24th, 26th, 29th to 36th, 39th, 44th, 45th, 48th
ORF2	the 3rd to 5th, 14th, 17th, 21st, 23rd, 24th, 28th, 31st, 34th, 36th, 37th, 41st, 44th, 46th, 49th, 50th
ORF3	the 2nd, 6th, 9th to 15th, 18th to 20th, 23rd to 26th, 28th to 31st, 33rd to 36th, 38th, 39th, 41st, 44th, 45th, 47th to 49th
ORF4	the 2nd to 10th, 12th, 13th, 15th to 18th, 20th, 22nd to 25th, 27th to 29th, 33rd, 34th, 36th, 37th, 39th, 40th, 43rd, 45th, 46th, 47th to 50th
ORF5	the 7th, 11th, 13th, 15th to 17th, 20th, 21st, 25th to 31st, 33rd, 37th, 39th, 41st, 50th
ORF6	the 11th, 17th, 18th, 20th, 23rd, 25th, 33rd, 38th, 41st, 43rd, 44th, 49th
ORF7	the 2nd, 5th, 6th, 8th, 9th, 15th, 18th, 21st to 23rd, 26th to 31st, 36th, 38th to 40th, 46th

listed in Table 4. Depending on the data from Table 3, 4 and comparison with the whole coding sequence of PRRSV, for ORF1a, the codons with negative CUB, namely GCA (Ala), GCG (Ala), CAA (Gln), AGU (Ser), ACA (Thr) and ACG (Thr), were more preferentially chosen in the target region for both serotypes; for ORF2, the codons, namely UGU (Cys), AUA (Ile), AAA (Lys), CCG (Pro), AGU (Ser) and UCG (Ser), were more preferentially used; for ORF3, the codons, namely UGU (Cys), AGC (Ser) and ACG (Thr), were more preferentially chosen; for ORF4, the codons, namely GAC (Asp), UUC (Phe), AGU (Ser) and UCG (Ser), were more preferentially chosen; for ORF5, the codons, namely UGU (Cys), CCG (Pro), UCG (Ser) and ACG (Thr), were more preferentially chosen; for ORF6, the codons, namely CAA (Gln), AUA (Ile) and CUA (Leu), were more preferentially used; for ORF7, the codons, namely GGA (Gly) and AAA (Lys), were more preferentially chosen. Due to these non-preferential codons, ribosomes might be stalled by them to regulate the efficiency of gene translation.

3.4. The characteristics of codon usage for the target positions

The positions with negative CUB do not always use the codons with negative CUB, and the R value for the codons with negative CUB vary compared with $R = 1.00$. However, some target positions contain the codons with negative CUB and R values > 1.00, suggesting that some new characteristics might influence the translation efficiency of the corresponding coding sequence. In translation initiation region of ORF1a, the non-preferential codons (R value > 1.00) are preferentially used in the 4th (US and EU serotypes), 9th (US), 12th (EU), 19th (US), the 22nd (US and EU), 27th (US), 31st (US and EU) and 40th (US), while some non-preferential codons, which have R value < 1.00 or R value > 1.00, exist in the 16th (EU) and 30th (US and EU) positions. For ORF2, the non-preferential codons are more preferentially used in the 7th (EU), 8th (EU), 9th (EU), 11th (US), 20th (EU), 27th (EU), 30th (US), 33rd (US), 40th (EU), 43rd (EU), 44th (US) and 48th (EU) positions, while some non-preferential codons with R value > 1.00 or R value < 1.00 exist in the 12th (US) position. For ORF3, non-preferential codons (R value >

1.00) exist in the 4th (US), 13th (US and EU), 17th (EU), 26th (US and EU), 31st (US and EU), 32nd (EU) and 37th (US) positions, while the non-preferential codons with R value > 1.00 or R value < 1.00 are used in the 5th (EU), 6th (US and EU), 7th (US), 11th (US and EU),16th (US) and 43rd (EU) positions. For ORF4, the non-preferential codons with R value > 1.00 are used in the 3rd (US and EU), 7th (US and EU), 20th (US and EU), 27th (US and EU), 28th (US and EU), 29th (US and EU), 38th (EU),40th (US and EU), 41st (US), 44th (EU) and 49th (US and EU), while some non-preferential nodons with R value > 1.00 or R value < 1.00 are used in the 31st (EU) position. For ORF5, the non-preferential nodons with R value > 1.00 are used in the 9th (EU), 12th (US), 14th (EU), 22nd (US) 23rd (US), 32nd (US), 36th (US), 39th (EU), 40th (EU), 44th (EU), 48th (EU) and 49th (EU), while non-preferential codon with R value > 1.00 or R value > 1.00 are used in the 8th (US), 24th (US), 46th (US) and 47th (US) positions. For ORF6, the non-preferential codons (R value > 1.00) are used in the 3rd (US), 4th (EU), 7th (US), 13th (US), 14th (EU), 15th (EU), 19th (EU), 21st (US), 22nd (EU), 24th (EU), 26th (EU), 27th (US), 30th (EU), 31st (EU), 32nd (US), 37th (US), 40th (US), 45th (EU) and 48th (EU) positions, while some non-preferential codon (R value < 1.00 or R value > 1.00) are used in the 2nd (EU), 5th (US and EU), 46th (US) and 50th (US) positions. For ORF7, the non-preferential codons (R value > 1.00) are chosen in the 11th (US), 32nd (US), 40th (US and EU), 41st (US), 43rd (EU), 44th (EU), 48th (US) and 50th (US) positions, while some non-preferential codon (R value > 1.00 or R value < 1.00) are used in the 3rd (US), 24th (EU), 25th (EU) and 35th (US) positions. The rest positions with negative CUB do not arise from the existence of non-preferential codons but contain some preferential codons (CUB > 0), implying that these positions do not affect the efficiency of gene translation. The degeneracy of the genetic code enables the same amino acid sequences to be encoded and translated in different ways. However, the synonymous codon usage is not purely random.

4. Discussion

RNA virus possesses high mutation rates and therefore virus populations exist as dynamic and complex mutant

Table 3 Preferentially used codons in the target region in US and EU serotypes of PRRSV

Codon	ORF1a		ORF2		ORF3	
	US	EU	US	EU	US	EU
[a]GCA	[b]1.69	[b]1.43	0	0.91	0.15	0.64
GCC	1.00	0.98	1.42	2.16	0	0.91
[a]GCG	[b]2.25	[b]1.34	0	0	[b]2.17	0
GCU	0	0.76	1.03	0.83	2.94	1.88
[a]AGA	0	0	0.5	[b]1.37	0	0
AGG	1.22	0	0	0	3.72	0
[a]CGA	[b]2.42	0	0	0	0	0
CGC	0	0	0	0	0	2.05
CGG	3.07	5.63	3.79	1.14	0	0
[a]CGU	0	0	0	[b]1.07	0.47	[b]2.90
AAC	0.43	0.90	1.86	0	0.12	1.01
[a]AAU	[b]1.72	0.68	0.40	0	[b]1.35	1.00
[a]GAC	0	[b]1.56	0	0	0	0
GAU	2.31	0.37	0	1.26	0.38	0.25
UGC	1.08	1.62	1.25	0	0.36	0.18
[a]UGU	0.92	0.50	[b]1.01	[b]1.51	[b]1.44	[b]1.43
[a]CAA	[b]1.89	[b]1.08	0.50	0.86	0	0
CAG	0.17	0.92	0	1.21	0	2.15
[a]GAA	0.55	0.28	0	0	[b]1.50	0
GAG	1.30	1.38	0	1.14	0.27	1.30
[a]GGA	0	0.39	0	[b]2.46	0	0
GGC	1.40	1.63	3.13	0	1.51	3.82
GGG	2.00	1.16	0	0.29	0	0
GGU	0.15	0.48	1.40	1.25	2.58	0.42
[a]CAC	0	0	0	[b]1.58	0	0.72
CAU	0	0	0	0.21	1.41	1.32
[a]AUA	[b]8.26	0	[b]2.55	[b]3.19	0	0
AUC	0	0	0	0	0	2.58
AUU	0	0.59	0	0	4.31	0
[a]CUA	[b]2.32	0.24	0.32	0	0	[b]4.42
CUC	1.90	1.87	0	0	1.84	0.77
CUG	0.92	0	0	0.65	1.39	0.43
CUU	1.01	1.34	0.66	0.24	0	1.02
[a]UUA	0.28	0.22	0.64	[b]3.54	0	0
UUG	0	0.80	1.99	2.04	0.74	1.29
[a]AAA	0	0	[b]1.27	[b]2.64	0	0
AAG	0	0	0	0	0	0
[a]UUC	[b]1.01	0.48	0.17	[b]1.09	0.99	0.95
UUU	1.01	1.46	1.14	0.78	1.02	1.07
CCA	0	0.98	1.50	1.04	0	4.51
CCC	2.27	0.11	0	0	0.39	0
[a]CCG	0	[b]2.43	[b]1.36	[b]1.07	[b]3.79	0
CCU	1.01	0.81	0	3.21	0	0.24
[a]AGC	0.42	[b]1.56	0	0.43	[b]2.02	[b]1.26
[a]AGU	[b]6.36	[b]8.28	[b]4.94	[b]4.13	0	[b]1.87
UCA	1.89	1.68	0.28	0.41	0	1.17
UCC	0.21	0.98	0.09	0.60	0.83	0.59
[a]UCG	0	0	[b]1.58	[b]1.70	0	[b]1.30
UCU	2.79	1.33	1.16	0.63	2.21	1.22

Table 3 Preferentially used codons in the target region in US and EU serotypes of PRRSV *(Continued)*

Codon	ORF1a		ORF2		ORF3	
[a]ACA	[b]1.97	[b]1.48	[b]6.17	0	[b]1.24	0.70
ACC	0.89	0.58	0	0	0	0.70
[a]ACG	[b]2.22	[b]1.49	0	[b]1.83	[b]1.74	[b]2.11
[a]ACU	0	[b]1.06	0	0	0.96	0.82
UAC	1.33	0	0	1.36	1.32	2.03
[a]UAU	0.37	0.30	[b]1.77	0.53	0	0.11
[a]GUA	0	[b]2.67	0	[b]4.79	0	0
GUC	0.77	1.77	3.78	0.33	0.11	0.33
GUG	1.71	0.31	0	2.04	1.88	0
GUU	0.76	0.53	2.96	0	1.08	4.14

Codon	ORF4		ORF5		ORF6		ORF7	
	US	EU	US	EU	US	EU	US	EU
[a]GCA	[b]1.25	0.16	0	0	0	[b]1.60	0	[b]1.09
GCC	0.55	0.96	1.91	3.08	0.65	1.69	1.22	1.75
[a]GCG	[b]1.27	[b]1.09	0.85	0.57	[b]1.60	0	0	0
GCU	0.98	1.20	1.35	0	2.24	0.21	1.50	0.74
[a]AGA	0	0	0	[b]3.69	0	0	[b]2.78	0.71
AGG	0	0	0	0.15	0	0	0.17	2.19
[a]CGA	0	0	[b]7.83	0	[b]4.58	0.99	0	0.33
CGC	0	0	0	0	2.29	5.79	0	0.75
CGG	0	0	0	0	0	0.23	0	0.29
[a]CGU	0	0	0	[b]1.80	0	0	0	0
AAC	3.67	1.36	1.34	1.42	0	0.31	1.09	1.42
[a]AAU	0	0.56	0	0.20	0.29	0.86	0.89	0.83
[a]GAC	[b]1.38	[b]1.39	0.58	0.18	[b]1.33	0.64	0.50	0
GAU	0.50	0.69	0	2.65	0.67	0.86	0.75	0
UGC	0.64	0.67	0.82	0.99	0.44	1.38	1.92	1.00
[a]UGU	[b]1.41	0.99	[b]1.13	[b]1.02	[b]1.33	0.24	0	0.29
[a]CAA	0.42	[b]1.14	0.25	1.00	[b]1.50	[b]1.38	0.64	0.78
CAG	1.46	0.95	2	0	0.33	0	1.13	1.07
[a]GAA	0.50	0	0	0	0.17	0.45	0	0
GAG	0.19	2.14	1.49	0.29	0	0	0	0
[a]GGA	0	0	0	0	0	0.84	[b]1.31	[b]1.14
GGC	0.63	0	1.63	1.98	0.79	0.77	1.05	0.89
GGG	0	0	1.03	0.54	1.87	1.97	0.63	0.85
GGU	4.00	3.33	0.17	0	0	0.56	1.35	1.11
[a]CAC	0	0	0	[b]2.90	0.78	[b]2.05	0	0
CAU	0	1.48	1.94	0	1.46	0	0	0
[a]AUA	0	0	0	[b]2.58	[b]1.12	[b]1.86	0	[b]3.29
AUC	1.26	1.24	1.07	0.12	1.19	0.90	1.46	0
AUU	0.73	1.20	1.34	0.37	0.80	0.08	0	0
[a]CUA	0	[b]2.26	[b]1.37	0	[b]1.66	[b]1.35	0	0
CUC	1.52	1.15	0.60	0	0.19	0.74	0	0
CUG	0	1.14	0.70	0.99	1.19	1.48	1.46	1.78
CUU	1.19	1.81	0.62	0.77	1.27	0.94	0	0
[a]UUA	0	0.23	0.55	0	0.77	0.16	0	0
UUG	1.50	0.19	1.69	2.16	0.73	0.84	0.22	1.05
[a]AAA	[b]1.21	0.20	[b]1.34	1.00	0	0.31	[b]1.08	[b]1.21
AAG	0.69	1.82	0.23	0.38	2.43	1.32	0.98	0.86
[a]UUC	[b]1.28	[b]1.18	[b]1.88	0.99	1.00	0.42	0	0
UUU	0.76	0.12	0.84	0.98	0.87	1.49	0	0

Table 3 Preferentially used codons in the target region in US and EU serotypes of PRRSV *(Continued)*

CCA	3.83	0	0	0.76	3.00	0	2.25	1.85
CCC	0.40	1.67	0	0	0	0.79	0	0.59
[a]CCG	0	0.43	[b]4.25	[b]1.29	0	0	[b]2.00	0.33
CCU	0	0.24	0.17	0	0	1.65	0	0.50
[a]AGC	[b]1.34	0	[b]1.69	0.57	0.98	[b]3.10	1.00	[b]2.47
[a]AGU	[b]3.67	[b]2.57	0	0	[b]4.28	0	0	[b]2.51
UCA	4.31	0.64	0	0	2.20	0	0	2.99
UCC	0.78	0	0.32	1.17	0.79	0	1.35	1.26
[a]UCG	[b]1.91	[b]1.95	[b]1.58	[b]1.48	[b]2.20	0	0	0
UCU	0.47	0.86	1.45	1.78	2.02	0.31	0	0
[a]ACA	0	0	0	[b]1.32	0.17	[b]1.76	0	[b]1.63
ACC	1.78	1.41	0.98	0.67	0.83	0.56	0	0
[a]ACG	0	[b]1.69	[b]6.79	[b]2.13	[b]2.35	0	0	0
[a]ACU	0.78	0.21	0	0.53	0.18	0.43	0	0
UAC	0	0	1.80	1.11	0.67	1.05	0	0
[a]UAU	0	0	0.48	0.90	[b]1.50	0.55	0	0
[a]GUA	0	0	0	0	0.98	0.27	0	0
GUC	0.92	1.04	0.34	0	0	0	1.72	2.96
GUG	0.52	0.10	2.79	0	2.10	2.29	0	0
GUU	2.25	1.77	0	0	0	0	0	0

[a] presented the non-preferential codon.
[b] presented that the non-preferential codon was more preferentially chosen in the translation initiation region than that of the whole coding sequence.

Table 4 Synonymous codon usage bias for the whole coding sequence of PRRSV

[a]AA	Codon	[b]CUB_{ij}	AA[a]	Codon	[b]CUB_{ij}
Ala	GCA	-0.143	Leu	CUA	-0.490
	GCC	0.309		CUC	0.164
	GCG	-0.350		CUG	0.377
	GCU	0.184		CUU	0.096
Arg	AGA	-0.107		UUA	-0.670
	AGG	0.092		UUG	0.522
	CGA	-0.258	Lys	AAA	-0.003
	CGC	0.411		AAG	0.003
	CGG	0.012	Phe	UUC	-0.083
	CGU	-0.149		UUU	0.083
Asn	AAC	0.040	Pro	CCA	0.047
	AAU	-0.04		CCC	0.059
Asp	GAC	-0.045		CCG	-0.198
	GAU	0.045		CCU	0.092
Cys	UGC	0.001	Ser	AGC	-0.128
	UGU	-0.001		AGU	-0.194
Gln	CAA	-0.018		UCA	0.075
	CAG	0.018		UCC	0.418
Glu	GAA	-0.116		UCG	-0.317
	GAG	0.116		UCU	0.145
Gly	GGA	-0.441	Thr	ACA	-0.031
	GGC	0.369		ACC	0.439
	GGG	0.016		ACG	-0.347
	GGU	0.056		ACU	-0.061
His	CAC	-0.153	Tyr	UAC	0.174
	CAU	0.153		UAU	-0.174
Ile	AUA	-0.234	Val	GUA	-0.635
	AUC	0.166		GUC	0.110
	AUU	0.068		GUG	0.406
				GUU	0.118

[a] mean amino acid
[b]The CUB_{ij} value was calculated following by the equation: $CUB_{ij} = RSCU_{ij} - RSCU_0$, and $RSCU_{ij}$ value came from the previous study [30].

distributions [36-41]. However, the redundant intensity of mutation has deleterious effects on the viral fitness. Thus, the robustness of viral sequences can perform a reduced sensitivity to perturbations affecting phenotypic expression. The balance between the high mutations and the robustness produce a dynamic population pool, termed as 'quasispecis' [36,42]. As to comparative genomics, it is generally accepted that sequences with a crucial function are conserved among different but related organisms [43-45]. In addition, Akashi found that the frequency of preferential codons is significantly higher at the conserved amino acid positions than that at the non-conserved amino acid positions among different *Drosophila* species, suggesting that translation selection favors the conserved pattern of synonymous codon usage to enhance the accuracy of gene expression [46]. A lot of experimental data have shown that rates of chain elongation during translation of proteins are not uniform [47]. Non-uniform character of distribution of codons with different usage frequencies along mRNA is assumed to be a main factor to modulate the translation rate. Extensive studies have been carried out previously on the determination of the translation rates and the overall level of gene expression for certain individual codons [48-52]. From this research, we observed that the conserved pattern of codon usage did not simply follow the corresponding positions in the conserved sequence fragment, suggesting that the conservation of codon usage within a gene sequence have an

important function in modulating its translational rate. The positions with the conserved positive CUB enhance the accuracy and efficiency of their gene translation. It has been observed that preferential codons can reduce the frequency of amino acid misincorporations, resulting in an approximately 10-fold increase of protein products over non-preferential codons for the same amino acid [53]. However, the positions with negative CUB in the translation initiation region of each PRRSV subgenomic RNA are not ignored. Because these positions are likely to regulate the translation initiation rate to generate the target product with high activity. Lithwich and Margalit reported that CUB is most highly associated with protein expression and is most conserved [26]. Once a significant number of gene sequences have been obtained, it will be taken into consideration that biased codon usage can regulate the expression levels of individual genes by modulating

the rates of polypeptide elongation [21,54-58]. Komar pointed out that although preferential codons enable the corresponding gene to be translated efficiently, the non-preferential codons replaced by the corresponding preferential codons can regulate the gene expression to perform the precise protein folding [59]. Lavner and Kotlar indicated that translation selection may shape codon bias pattern, not only to increase translation efficiency by favoring preferential codons in highly expressed genes, but also to decrease translation rate by favoring non-optimal codons in lowly expressed ones [60]. A relationship between the translation efficiency and CUB have been reported that it can lead to link between the protein folding by modulating the translational rate and the synonymous codon usage bias [47,61-65]. The nucleotide sequences around the N-terminal region of the protein appear to be particularly sensitive to the presence of rare codons [66,67]. Our data showed that some positions in the translation initiation regions of ORFs tended to preferentially choose non-preferential codons which were more preferentially used in these regions than the whole coding sequences. This phenomenon suggested that the determinant of the invariant pattern of codon usage is not only correlated with the conserved sequence, but also dependent of the translation selection. As codon usage pattern comprised of preferential and non-preferential codons contributes to different translation rates, it is possible to change the local translation rates of a gene by suitable selection of its synonymous codons. A gene sequence with non-preferential codons intends to encode turns, loops and domain linkers within its protein structure through the limited step to the translation rate [47,63,64,68]. Taken together, under the translation selection, the conserved non-preferential codons in the translation initiation regions of PRRSV may affect the translation efficiency so as to maintain the normal biological functions of their target products. Komar and Jaenicke indicated that the non-preferential coodns play an important role in maintaining the normal function or activity of *CAT* product [68]. It shows the importance of non-preferential codons to the formation of the target products. As non-preferential codons or even one aggregating near the translation initiation codon can decrease translation rate arising from the limitation of availability of tRNAs depending on the host cell [69], the view that non-preferential codons probably have a negative effect on gene expression can be explained by the 'minor codon modulator hypothesis' [70]. When the tRNA concentration of minor codons becomes extremely limited, ribosomes of the host cell block at the minor codons to inhibit the ribosome from entering into the initiation site effectively, thereby resulting in a decrease in the translation rate. Moreover, the non-preferential codons locating at the translation initiation region modulate the number of ribosomes that are sequestered by an mRNA if the rates of

elongation at these codons were so sufficiently slow that stalled ribosomes could block access to the initiation signals [19,71].

In summary, the conserved non-preferential codons in the translation initiation region have a high relationship with the regulation of gene expression. And the conserved codons with negative CUB are preferentially used in the initial region, which may be explained by the minor codon modulator hypothesis and the translation selection. These codons within this critical region might play a negative role in regulation of gene expression.

List of abbreviations
PRRSV: Porcine reproductive and respiratory syndrome virus; SCUV: synonymous codon usage values; CUB: codon usage bias; US: Northern American isolate; EU: European isolate.

Author details
[1]State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046, PR China. [2]College of Life Science and Engineering, Northwest University for Nationalities, Lanzhou, 730030, PR China. [3]The School of Public Health, Lanzhou University, Lanzhou, 730000, RP China. [4]School of Agriculture, Ningxia University, Yinchuan, 750021, RP China.

Authors' contributions
JHS and XXM carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. YLH, JDL and XSM participated in the sequence alignment. YXD and XNL participated in the design of the study and performed the statistical analysis. XPC conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Loula T: **Mystery pig disease.** *Agri-Practice (USA)* 1991.
2. Keffaber K: **Reproductive failure of unknown etiology.** *Am Assoc Swine Pract Newsl* 1989, **1**:1-9.
3. Bautista EM, Goyal SM, Yoon IJ, Joo HS, Collins JE: **Comparison of porcine alveolar macrophages and CL 2621 for the detection of porcine reproductive and respiratory syndrome (PRRS) virus and anti-PRRS antibody.** *Journal of Veterinary Diagnostic Investigation* 1993, **5**:163.
4. Collins JE, Benfield DA, Christianson WT, Harris L, Hennings JC, Shaw DP, Goyal SM, McCullough S, Morrison RB, Joo HS, *et al*: **Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs.** *J Vet Diagn Invest* 1992, **4**:117-126.
5. Meng XJ, Paul P, Halbur P, Lum M: **Phylogenetic analyses of the putative M (ORF 6) and N (ORF 7) genes of porcine reproductive and respiratory syndrome virus (PRRSV): implication for the existence of two genotypes of PRRSV in the USA and Europe.** *Archives of virology* 1995, **140**:745-755.
6. Nelsen CJ, Murtaugh MP, Faaberg KS: **Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents.** *J Virol* 1999, **73**:270-280.

7. Wensvoort G, Terpstra C, Pol J, Ter Laak E, Bloemraad M, De Kluyver E, Kragten C, Van Buiten L, Den Besten A, Wagenaar F: **Mystery swine disease in The Netherlands: the isolation of Lelystad virus.** *The Veterinary Quarterly* 1991, **13**:121.
8. Benfield DA, Nelson E, Collins JE, Harris L, Goyal SM, Robison D, Christianson WT, Morrison RB, Gorcyca D, Chladek D: **Characterization of swine infertility and respiratory syndrome (SIRS) virus (isolate ATCC VR-2332).** *Journal of Veterinary Diagnostic Investigation* 1992, **4**:127.
9. Cavanagh D: **Nidovirales: a new order comprising Coronaviridae and Arteriviridae.** *Arch Virol* 1997, **142**:629-633.
10. Conzelmann KK, Visser N, Van Woensel P, Thiel HJ: **Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group.** *Virology* 1993, **193**:329-339.
11. Meulenberg JJM, Hulst MM, de Meijer EJ, Moonen PLJM, den Besten A, de Kluyver EP, Wensvoort G, Moormann RJM: **Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV.** *Virology* 1993, **192**:62-72.
12. Snijder EJ, Meulenberg J: **The molecular biology of arteriviruses.** *Journal of general virology* 1998, **79**:961.
13. Spilman MS, Welbon C, Nelson E, Dokland T: **Cryo-electron tomography of porcine reproductive and respiratory syndrome virus: organization of the nucleocapsid.** *Journal of general virology* 2009, **90**:527.
14. Pasternak AO, Spaan WJM, Snijder EJ: **Nidovirus transcription: how to make sense...?** *Journal of general virology* 2006, **87**:1403.
15. Kim JK, Hollingsworth MJ: **Localization of in vivo ribosome pause sites.** *Analytical biochemistry* 1992, **206**:183-188.
16. Miyasaka H: **The positive relationship between codon usage bias and translation initiation AUG context in Saccharomyces cerevisiae.** *Yeast* 1999, **15**:633-637.
17. MIYASAKA H, KANAI S, TANAKA S, AKIYAMA H, HIRANO M: **Statistical analysis of the relationship between translation initiation AUG context and gene expression level in humans.** *Bioscience, biotechnology, and biochemistry* 2002, **66**:667-669.
18. Stanssens P, Remaut E, Fiers W: **Inefficient translation initiation causes premature transcription termination in the lacZ gene.** *Cell* 1986, **44**:711-718.
19. Zhou J, Zhang J, Ding Y, Chen H, Ma L, Liu Y: **Characteristics of codon usage bias in two regions downstream of the initiation codons of foot-and-mouth disease virus.** *Biosystems* 2010, **101**:20-28.
20. Ohno H, Sakai H, Washio T, Tomita M: **Preferential usage of some minor codons in bacteria.** *Gene* 2001, **276**:107-115.
21. Gouy M, Gautier C: **Codon usage in bacteria: correlation with gene expressivity.** *Nucleic acids research* 1982, **10**:7055.
22. Grantham R, Gautier C, Gouy M, Mercier R, Pave A: **Codon catalog usage and the genome hypothesis.** *Nucleic Acids Res* 1980, **8**:r49-r62.
23. Gustafsson C, Govindarajan S, Minshull J: **Codon bias and heterologous protein expression.** *Trends Biotechnol* 2004, **22**:346-353.
24. Kudla G, Murray AW, Tollervey D, Plotkin JB: **Coding-sequence determinants of gene expression in Escherichia coli.** *Science* 2009, **324**:255-258.
25. Liljenstrom H, von Heijne G: **Translation rate modification by preferential codon usage: intragenic position effects.** *J Theor Biol* 1987, **124**:43-55.
26. Lithwick G, Margalit H: **Hierarchy of sequence-dependent features associated with prokaryotic translation.** *Genome Res* 2003, **13**:2665-2673.
27. Post LE, Nomura M: **DNA sequences from the str operon of Escherichia coli.** *J Biol Chem* 1980, **255**:4660-4666.
28. Sharp PM, Emery LR, Zeng K: **Forces that influence the evolution of codon bias.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2010, **365**:1203.
29. Vicario S, Moriyama EN, Powell JR: **Codon usage in twelve species of Drosophila.** *BMC evolutionary biology* 2007, **7**:226.
30. Liu YS, Zhou JH, Chen HT, Ma LN, Ding YZ, Wang M, Zhang J: **Analysis of synonymous codon usage in porcine reproductive and respiratory syndrome virus.** *Infect Genet Evol* 2010, **10**:797-803.
31. Jacques N, Dreyfus M: **Translation initiation in Escherichia coli: old and new questions.** *Mol Microbiol* 1990, **4**:1063-1067.
32. Braiman M, Stern LJ, Chao BH, Khorana H: **Structure-function studies on bacteriorhodopsin. IV. Purification and renaturation of bacterio-opsin polypeptide expressed in Escherichia coli.** *Journal of Biological Chemistry* 1987, **262**:9271.

33. Andersson SG, Kurland CG: **Codon preferences in free-living microorganisms.** *Microbiol Rev* 1990, **54**:198-210.
34. Hooper SD, Berg OG: **Gradients in nucleotide and codon usage along Escherichia coli genes.** *Nucleic acids research* 2000, **28**:3517.
35. Liu YS, Zhou JH, Chen HT, Ma LN, Pejsak Z, Ding YZ, Zhang J: **The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern.** *Infect Genet Evol* 2011, **11**:1168-1173.
36. Domingo E, Holland JJ: **RNA virus mutations and fitness for survival.** *Annu Rev Microbiol* 1997, **51**:151-178.
37. Elena SF, Miralles R, Cuevas JM, Turner PE, Moya A: **The two faces of mutation: extinction and adaptation in RNA viruses.** *IUBMB Life* 2000, **49**:5-9.
38. Elena SF: **Restrictions to RNA virus adaptation: an experimental approach.** *Antonie Van Leeuwenhoek* 2002, **81**:135-142.
39. Elena SF, Sanjuan R: **RNA viruses as complex adaptive systems.** *Biosystems* 2005, **81**:31-41.
40. Elena SF, Carrasco P, Daros JA, Sanjuan R: **Mechanisms of genetic robustness in RNA viruses.** *EMBO Rep* 2006, **7**:168-173.
41. Klein J: **Understanding the molecular epidemiology of foot-and-mouth-disease virus.** *Infect Genet Evol* 2009, **9**:153-161.
42. Wilke CO: **Quasispecies theory in the context of population genetics.** *BMC evolutionary biology* 2005, **5**:44.
43. Carrillo C, Tulman ER, Delhon G, Lu Z, Carreno A, Vagnozzi A, Kutish GF, Rock DL: **Comparative genomics of foot-and-mouth disease virus.** *J Virol* 2005, **79**:6487-6504.
44. Chen Z, Li K, Plagemann PGW: **Neuropathogenicity and sensitivity to antibody neutralization of lactate dehydrogenase-elevating virus are determined by polylactosaminoglycan chains on the primary envelope glycoprotein.** *Virology* 2000, **266**:88-98.
45. Han J, Rutherford MS, Faaberg KS: **The porcine reproductive and respiratory syndrome virus nsp2 cysteine protease domain possesses both trans-and cis-cleavage activities.** *Journal of virology* 2009, **83**:9449.
46. Akashi H: **Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy.** *Genetics* 1994, **136**:927-935.
47. Krasheninnikov IA, Komar AA, Adzhubei IA: **Nonuniform size distribution of nascent globin peptides, evidence for pause localization sites, and a contranslational protein-folding model.** *J Protein Chem* 1991, **10**:445-453.
48. Curran JF, Yarus M: **Rates of aminoacyl-tRNA selection at 29 sense codons in vivo* 1.** *Journal of molecular biology* 1989, **209**:65-77.
49. S rensen MA, Pedersen S: **Absolute in vivo translation rates of individual codons in Escherichia coli* 1:: The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate.** *Journal of molecular biology* 1991, **222**:265-280.
50. Martin SL, Vrhovski B, Weiss AS: **Total synthesis and expression in Escherichia coli of a gene encoding human tropoelastin.** *Gene* 1995, **154**:159-166.
51. Mohsen AW, Vockley J: **High-level expression of an altered cDNA encoding human isovaleryl-CoA dehydrogenase in Escherichia coli.** *Gene* 1995, **160**:263-267.
52. Nakamura T, Suyama A, Wada A: **Two types of linkage between codon usage and gene-expression levels.** *FEBS letters* 1991, **289**:123-125.
53. Precup J, Parker J: **Missense misreading of asparagine codons as a function of codon identity and context.** *Journal of Biological Chemistry* 1987, **262**:11351.
54. Chavancy G, Garel JP: **Does quantitative tRNA adaptation to codon content in mRNA optimize the ribosomal translation efficiency? Proposal for a translation system model.** *Biochimie* 1981, **63**:187-195.
55. Bonekamp F, Andersen HD, Christensen T, Jensen KF: **Codon-defined ribosomal pausing in Escherichia coli detected by using the pyrE attenuator to probe the coupling between transcription and translation.** *Nucleic acids research* 1985, **13**:4113.
56. Grosjean H, Fiers W: **Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes.** *Gene* 1982, **18**:199-209.
57. Robinson M, Lilley R, Little S, Emtage J, Yarranton G, Stephens P, Millican A, Eaton M, Humphreys G: **Codon usage can affect efficiency of translation of genes in Escherichia coli.** *Nucleic acids research* 1984, **12**:6663.
58. Maroun LE, Degner M, Precup JW, Franciskovich PP: **Eukaryotic mRNA 5'-Leader sequences have dual regions of complementarity to the 3'-terminus of 18s rRNA*.** *Journal of theoretical biology* 1986, **120**:85-98.

59. Komar AA, Lesnik T, Reiss C: **Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation.** *FEBS Lett* 1999, **462**:387-391.
60. Lavner Y, Kotlar D: **Codon bias as a factor in regulating expression via translation rate in the human genome.** *Gene* 2005, **345**:127-138.
61. Adzhubei IA, Adzhubei AA, Neidle S: **An Integrated Sequence-Structure Database incorporating matching mRNA sequence, amino acid sequence and protein three-dimensional structure data.** *Nucleic Acids Res* 1998, **26**:327-331.
62. Oresic M, Shalloway D: **Specific correlations between relative synonymous codon usage and protein secondary structure1.** *Journal of molecular biology* 1998, **281**:31-48.
63. Purvis IJ, Bettany AJE, Santiago TC, Coggins JR, Duncan K, Eason R, Brown AJP: **The efficiency of folding of some proteins is increased by controlled rates of translation in vivo:: A hypothesis.** *Journal of molecular biology* 1987, **193**:413-417.
64. Thanaraj T, Argos P: **Ribosome-mediated translational pause and protein domain organization.** *Protein Science: A Publication of the Protein Society* 1996, **5**:1594.
65. Tao X, Dafu D: **The relationship between synonymous codon usage and protein structure.** *FEBS letters* 1998, **434**:93-96.
66. Deana A, Ehrlich R, Reiss C: **Silent mutations in the Escherichia coli ompA leader peptide region strongly affect transcription and translation in vivo.** *Nucleic Acids Res* 1998, **26**:4778-4782.
67. Hoekema A, Kastelein RA, Vasser M, de Boer HA: **Codon replacement in the PGK1 gene of Saccharomyces cerevisiae: experimental approach to study the role of biased codon usage in gene expression.** *Mol Cell Biol* 1987, **7**:2914-2924.
68. Komar AA, Jaenicke R: **Kinetics of translation of [gamma] B crystallin and its circularly permutated variant in an in vitro cell-free system: possible relations to codon distribution and protein folding.** *FEBS letters* 1995, **376**:195-198.
69. Chen GT, Inouye M: **Role of the AGA/AGG codons, the rarest codons in global gene expression in Escherichia coli.** *Genes Dev* 1994, **8**:2641-2652.
70. Chen GFT, Inouye M: **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 research* 1990, **18**:1465.
71. Belsham G: **Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes in vivo.** *The EMBO Journal* 1992, **11**:1105.