

## Research Article

# The Effects of the Context-Dependent Codon Usage Bias on the Structure of the *nsp1 $\alpha$* of Porcine Reproductive and Respiratory Syndrome Virus

Yao-zhong Ding,<sup>1,2</sup> Ya-nan You,<sup>1</sup> Dong-jie Sun,<sup>1</sup> Hao-tai Chen,<sup>1,2</sup> Yong-lu Wang,<sup>1,2</sup> Hui-yun Chang,<sup>1,2</sup> Li Pan,<sup>1,2</sup> Yu-zhen Fang,<sup>1,2</sup> Zhong-wang Zhang,<sup>1,2</sup> Peng Zhou,<sup>1,2</sup> Jian-liang Lv,<sup>1,2</sup> Xin-sheng Liu,<sup>1,2</sup> Jun-jun Shao,<sup>1,2</sup> Fu-rong Zhao,<sup>1,2</sup> Tong Lin,<sup>1,2</sup> Laszlo Stipkovits,<sup>3</sup> Zygmunt Pejsak,<sup>4</sup> Yong-guang Zhang,<sup>1,2</sup> and Jie Zhang<sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Veterinary Etiological Biology, National Foot-and-Mouth Disease Reference Laboratory, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, Gansu 730046, China

<sup>2</sup> Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, Jiangsu 225009, China

<sup>3</sup> RT-Europe Research Center Ltd., Budapest, Hungary

<sup>4</sup> Department of Swine Diseases, National Veterinary Research Institute, 57 Partyzantow, 24-100 Pulawy, Poland

Correspondence should be addressed to Yong-guang Zhang; zhangyongguang@caas.cn and Jie Zhang; zhangjie03@caas.cn

Received 18 March 2014; Revised 5 June 2014; Accepted 19 June 2014; Published 3 August 2014

Academic Editor: Hongwei Wang

Copyright © 2014 Yao-zhong Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The information about the crystal structure of porcine reproductive and respiratory syndrome virus (PRRSV) leader protease *nsp1 $\alpha$*  is available to analyze the roles of tRNA abundance of pigs and codon usage of the *nsp1 $\alpha$*  gene in the formation of this protease. The effects of tRNA abundance of the pigs and the synonymous codon usage and the context-dependent codon bias (CDCB) of the *nsp1 $\alpha$*  on shaping the specific folding units ( $\alpha$ -helix,  $\beta$ -strand, and the coil) in the *nsp1 $\alpha$*  were analyzed based on the structural information about this protease from protein data bank (PDB: 3IFU) and the *nsp1 $\alpha$*  of the 191 PRRSV strains. By mapping the overall tRNA abundance along the *nsp1 $\alpha$* , we found that there is no link between the fluctuation of the overall tRNA abundance and the specific folding units in the *nsp1 $\alpha$* , and the low translation speed of ribosome caused by the tRNA abundance exists in the *nsp1 $\alpha$* . The strong correlation between some synonymous codon usage and the specific folding units in the *nsp1 $\alpha$*  was found, and the phenomenon of CDCB exists in the specific folding units of the *nsp1 $\alpha$* . These findings provide an insight into the roles of the synonymous codon usage and CDCB in the formation of PRRSV *nsp1 $\alpha$*  structure.

## 1. Introduction

Porcine reproductive and respiratory syndrome virus (PRRSV) is an economically important pathogen of swine. The PRRSV belongs to the order Nidovirales, family Arteriviridae, genus *Arterivirus* [1]. The PRRSV genome contains at least 9 open reading frames, including ORF1a encoding papain-like cysteine protease, ORF1b encoding RNA dependent RNA polymerase, ORFs 2–6 encoding envelop proteins, and ORF7 encoding the nucleocapsid protein [2, 3]. PRRSV strains can be divided into two distinct serotypes,

namely, the North American isolate (US) and the European isolate (EU) [4–8].

The replicative enzymes of the PRRSV are encoded in ORF1a and ORF1b, which locate in the 5' proximal three quarters of the viral genome. The two polyproteins encoded by ORF1a and ORF1b are cleaved extensively by the non-structural protein 4 (*nsp4*) deriving from ORF1a, yielding a series of nonstructural proteins [9]. In particular, the *nsp1* and the *nsp2* proteases release themselves from the ORF1a polyprotein firstly, and the *nsp1* can be further processed into two multifunctional proteases, namely, the *nsp1 $\alpha$*  and

the *nsplβ* [10, 11]. The arterivirus *nspl* region contains a tandem of papain-like autoprotease domains (PCP $\alpha$  and PCP $\beta$ ), and the arterivirus PCP $\alpha$  and PCP $\beta$  domains were found to be active in the reticulocyte lysates and the *E. coli* systems [12, 13]. This biological feature might indicate that the active functions of PCP $\alpha$  and PCP $\beta$  are free from the different types of the expression systems and depend on the correct folding by themselves. As for the *nspl $\alpha$* , it plays an important role in regulating the accumulation of both genome- and subgenome-length minus-strand RNA and thereby fine-tuning the relative abundance of each of viral mRNAs in the infected cells [10, 14, 15]. The correct secondary structure of the *nspl $\alpha$*  is required for the biological functions of the protease. Based on the crystal structure of the *nspl $\alpha$* , it was found that this nonstructural protein has three domains, namely, the N-terminal zinc finger (ZF) domain, the papain-like cysteine protease domain, and the carboxyl-terminal extension [16]. Recently, the role of the *nspl $\alpha$*  in impairing the host immune response has been reported [17]; however, little information about the relationship between synonymous codon usage and the secondary structure of the PRRSV *nspl $\alpha$*  is available to date.

The synonymous codon usage and translational speed of gene play important roles in many biological functions, like translation efficiency, genetic diversity, amino acid conservation, transfer RNA abundance, coevolution of the virus and its hosts, and context-dependent codon bias (CDCB), and so forth [18–22]. The nucleotide composition of a coding sequence (CDS) is nonrandom, and the CDS nonrandomness is influenced by the preferences in the selection of synonymous codons pairing to the same amino acid (termed as the synonymous codon usage bias SCUB). The link between SCUB and specific folding unit of protein gives us a new insight into the correct formation of the secondary structure of proteins [23–26]. It is noted that mRNA sequences generally have an additional potential to carry correct structural information in the forms of SCUB, which can be involved in a single codon or a nucleotide context of the target coding sequence [27, 28]. As for SCUB, neighboring nucleotides flanking a codon regulate the usage of the specific codon from the synonymous family, termed as context-dependent codon bias (CDCB) [20, 29–31]. It has been reported that the most important nucleotide determining CDCB is the first nucleotide after a codon, termed as the  $N_1$  context [32]. Although several evidences indicate the link between SCUB and the formation of the specific folding unit of viral protein, little information about the role of CDCB in the formation of the specific folding unit is reported up to date. In this study, we employed the structural information about the *nspl $\alpha$*  of PRRSV and several simple formulas to analyze the relationship between the CDCB of the PRRSV *nspl $\alpha$*  gene and the protease.

## 2. Materials and Methods

**2.1. Information of PRRSV Gene and Structure of the *nspl $\alpha$* .** The 191 coding sequences of PRRSV containing the *nspl $\alpha$*  gene were downloaded from the National Center

for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>) and the accession numbers of the sequences were listed in Table S1 available online at <http://dx.doi.org/10.1155/2014/765320>. To investigate SCUB of the *nspl $\alpha$* , the related genes were obtained from these 191 coding sequences by the multiple sequence alignments performed with the Clustal W (1.7) computer programs [33]. The information about the secondary structure of the PRRSV *nspl $\alpha$*  was obtained from protein data bank (PDB: 3IFU).

**2.2. Analysis of the Overall tRNA Abundance of Each Codon Position along the *nspl $\alpha$*  Gene.** To identify the translation selection caused by the various tRNA copy numbers (reflecting tRNA abundance) of the pigs (<http://gtrnadb.ucsc.edu/>) at each codon position in the PRRSV *nspl $\alpha$* , we devised an index ( $C$  value) representing the overall tRNA abundance for a particular codon position in a target gene. Consider

$$C = \sqrt[n]{\prod_{i=1}^n \left( \frac{W_{ij}}{W_j} \right)}, \quad (1)$$

where  $C$  value indicates the overall tRNA abundance for a particular codon position in the target gene,  $W_{ij}$  represents the tRNA copy numbers of a synonymous codon ( $i$ ) for the corresponding amino acid ( $j$ ),  $W_j$  represents the optimal tRNA copy numbers of a synonymous codon for the same amino acid, and  $n$  means the number of the interesting gene. The  $C$  value ranges from 0 to 1.0. The  $C$  value less than 0.3 for a codon position represents low tRNA abundance, and the  $C$  value more than 0.7 for a codon position represents high tRNA abundance.

**2.3. Estimation of the Relationship between the Synonymous Codon Usage Bias and the Secondary Structure of the *nspl $\alpha$* .** Based on the alignment between the amino acid sequences of the PRRSV (PDB: 3IFU) and the 191 *nspl $\alpha$*  genes involved in this study, we can locate the different folding units in the target protein. We devised the formula for the  $P$  value based on the previous research which analyzed the relationship between the codon usage bias and the structure of the target protein [25]. Consider

$$P = \ln \frac{f_{\text{obs}}}{f_{\text{exp}}},$$

$$f_{\text{obs}} = \frac{N_{(i,\text{sec}-k)}}{N_{(k)}}, \quad (2)$$

$$f_{\text{exp}} = \frac{\sum N_{(i,\text{sec}-j)}}{N_{\text{total}}},$$

where  $N_{(i,\text{sec}-k)}$  represents the amount of a specific synonymous codon for the corresponding amino acid in a specific folding unit (the  $\alpha$ -helix, the  $\beta$ -strand, or the coil) of protein;  $\text{sec}-k$  represents the corresponding amino acid in a specific secondary unit;  $N_{(k)}$  represents the amount of the amino acid in the corresponding folding unit. In addition,  $\sum N_{(i,\text{sec}-j)}$  represents the total number of amino acids in a specific

folding unit;  $sec-j$  contains the three kinds of folding unit, namely,  $\alpha$ -helix,  $\beta$ -strand, and the coil;  $N_{total}$  represents the total number of codons in the target genes. When the  $P$  value is more than zero, the corresponding synonymous codon ( $i$ ) owns a potential to be selected in a specific folding unit. When the  $P$  value is less than zero, the synonymous codon ( $i$ ) has no tendency to be chosen in a specific folding unit. Furthermore, we defined that when the  $P$  value is more than 0.1, the synonymous codon has a strong ability to exist in the specific folding unit; on the contrary, when the  $P$  value is less than  $-0.1$ , the synonymous codon has a strong tendency to avoid the specific folding unit.

**2.4. Calculation of the Relative Abundance of Codons with Context.** With the purpose to estimate the synonymous codons playing an important role in the formation of the specific folding units, codons having a significant tendency to exist in the specific folding unit of the PRRSV *nsplα* were analyzed by the formula for the relative abundance of codons with context. Berg and Silva [32] defined that the context  $N_1$  represents the first nucleotide after the target codon. Following this notation, we defined that the context  ${}_1N$  represents the last nucleotide before the target codon. We devised a formula calculating  $R$  value for the context  $N_1(xyz \sim n)$  and the context  ${}_1N(n \sim xyz)$  depending on the formula previously reported [20, 34]. Consider

$$R(xyz \sim n) = \frac{F(xyz \sim n)}{F(xyz)F(n)}, \tag{3}$$

$$R(n \sim xyz) = \frac{F(n \sim xyz)}{F(xyz)F(n)},$$

where  $F(xyz)$  is the frequency of the codon  $xyz$  and  $F(n)$  is the frequency of nucleotide  $n$  in the  $N_1$  or  ${}_1N$  context.  $F(xyz \sim n)$  and  $F(n \sim xyz)$  are the frequency of a codon with the  $n$  context. It is noted that  $x, y, z,$  and  $n$  are the nucleotides ( $a, u, g,$  or  $c$ ) and the codon is composed of  $xyz$ . Here and elsewhere the tilde character ( $\sim$ ) separates codons (italic) or oligonucleotides (nonunderlined) from their mononucleotide context.

**2.5. Calculation of the Relative Abundance of Mononucleotide and Dinucleotides in the *nsplα* Gene.** To investigate whether the  $N_1$  and  ${}_1N$  contexts are shaped by randomness or not, we calculated the frequencies of each nucleotide  $F(n)$  and dinucleotide  $F(xy)$ , where  $n, x, y,$  and  $z$  are each one of the four nucleotides ( $a, u, c,$  and  $g$ ). Then we calculated the relative abundances ( $r$  value) of the mononucleotide and dinucleotides with a single nucleotide context:  $r(n \sim x) = F(n \sim x)/[F(n)F(x)]$ , for mononucleotide  $x$  with context  $n$ ;  $r(xy \sim n) = F(xy \sim n)/[F(xy)F(n)]$ , for dinucleotide  $xy$  with context  $n$ .

**2.6. Statistic Analysis.** One-way analysis of variance, namely, one-way ANOVA, is a technique used to compare means of two or more samples. In this study, the ANOVA test is applied for identifying whether the overall tRNA abundance

of positions of a specific folding unit is different from other specific folding units or not. In addition, the ANOVA test is also employed to estimate whether the frequencies of codon usage in a specific folding unit are different from other specific folding units or not. This statistic analysis is carried out by the software SPSS 11.5.

### 3. Results

**3.1. The Overall tRNA Abundance for Each Codon Position of the *nsplα* Gene.** Based on the  $C$  values, the tRNA abundance for each codon position along the PRRSV *nsplα* gene was mapped. The translation speed for the synthesis of the *nsplα* is not stable in the pigs (Figure 1). The codon positions with the  $C$  values much less than 0.30 have a tendency to cluster in *nsplα* gene, including the positions 4–6, 8–10, 22–25, 27–30, 32–34, 38–40, 42–47, 50–53, 55–58, 68–70, 77–79, 83–85, 110–112, 119–122, 126–128, 139–141, 157–160, and 171–173. However, the codon positions with the  $C$  values much greater than 0.70 have few chances to cluster in *nsplα* gene which is translated in the pigs. Due to most codon positions with  $C$  values much less than 0.70 existing in the target gene, these positions within the *nsplα* might reduce the translation rate of this protein when the *nsplα* was scanned by the ribosomes in pig cells. It is noted that there are no significant differences ( $P > 0.05$ ) of the overall tRNA abundance for the codon positions in the regions of the three specific folding units of the *nsplα*. This result suggests that the fluctuation of the overall tRNA abundance pairing to each codon position along this *nsplα* might not regulate the formation of the specific folding units but decrease the scanning speed of ribosomes in the pig cells.

**3.2. The Relationship between the Synonymous Codon Usage Bias and the Structure of the *nsplα*.** Based on the  $P$  values for the synonymous codons which are involved in the formation of the specific folding units in the *nsplα*, we found the link between SCUD and the specific folding unit ( $P = 2.75 \times 10^{-11}$ ). In detail, the synonymous codons have a strong propensity toward shaping the  $\alpha$ -helix unit, including AUC for Ile, GUA for Val, AGC for Ser, AAG for Lys, and AUG for Met (Table 1). Turning to the effects of SCUB on shaping the  $\beta$ -strand unit, there are UUA for Leu, AUA for Ile, GUG for Val, UCA and AGU for Ser, ACA for Thr, UAC for Tyr, CGC for Arg, and two synonymous codons for His (Table 1). It is interesting that there are no codons which have a strong tendency to exist in the coil of the *nsplα* (Table 1). As for the codons which have a strong tendency ( $P$  value  $> 1.0$ ) to exist in the *nsplα*, all of them strongly tend to exist only in the  $\alpha$ -helix or the  $\beta$ -strand of this protein.

**3.3. The Relative Abundance of the Codon with  $N_1$  Context in the *nsplα* Gene.** As for the codons which have a strong tendency to exist in the specific folding unit of the *nsplα*, their  $R$  values, the relative abundance of codons with  $N_1$  contexts, were calculated from the 191 *nsplα* genes (Table 2). The data show that the occurrence of the codon with  $N_1$  context or  ${}_1N$  context is not random, and many codons with  $N_1$  context or  ${}_1N$  context have a strong tendency to

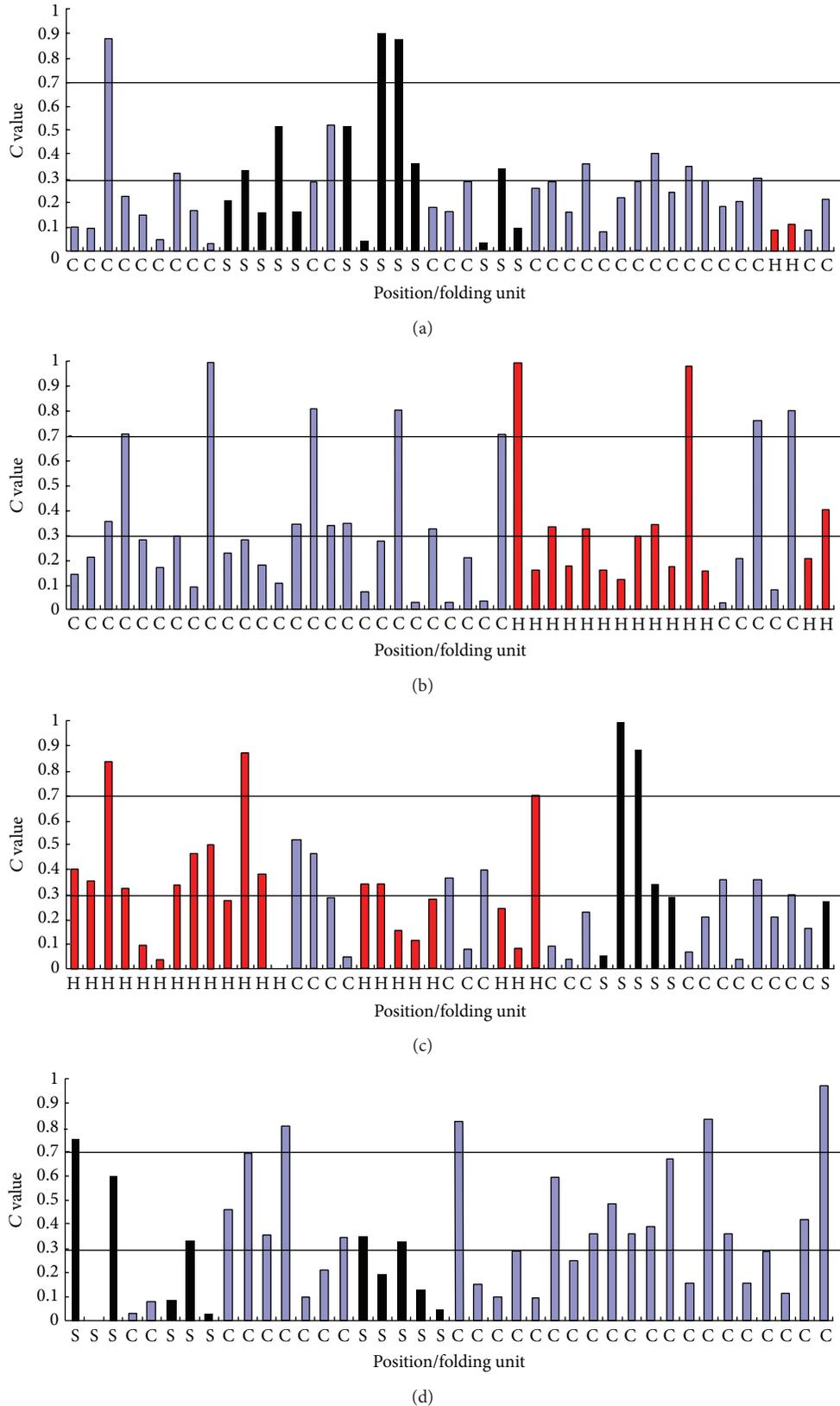


FIGURE 1: The overall tRNA abundance for each codon position of the PRRSV *nsplα* gene. The black bar corresponds to the  $\beta$ -strand region in the *nsplα*; the blue bar corresponds to the coil region in the *nsplα*; the red bar corresponds to helix region in the *nsplα*. (a) The codon positions range from the first position to the 45th position of the *nsplα*. (b) The codon positions range from the 46th position to the 90th position of the *nsplα*. (c) The codon positions range from the 91th position of the *nsplα*. (d) The codon positions range from the 136th to the 175th position of the *nsplα*.

TABLE 1: The relationship between the synonymous codon usage and the formation of the specific secondary structure unit.

Amino acid	Codon	<sup>a</sup> P value	<sup>b</sup> P value	<sup>c</sup> P value
Phe	UUU	0.40	0.07	-0.21
Phe	UUC	-0.92	-0.40	0.26
Leu	UUA	-1.07	<b>1.47</b>	-1.22
Leu	UUG	-0.16	-0.53	0.15
Leu	CUU	0.68	<sup>d</sup> —	-0.06
Leu	CUC	-5.53	-5.32	0.48
Leu	CUA	0.17	-1.37	0.13
Leu	CUG	0.00	-4.15	0.24
Ile	AUU	0.99	0.92	-5.46
Ile	AUC	<b>1.50</b>	-1.17	<sup>d</sup> —
Ile	AUA	0.69	<b>1.12</b>	-2.46
Val	GUU	0.58	0.28	-0.45
Val	GUC	0.32	0.20	-0.21
Val	GUA	<b>1.27</b>	0.02	-2.17
Val	GUG	-0.52	<b>1.21</b>	-0.73
Ser	UCU	0.76	<sup>d</sup> —	-0.12
Ser	UCC	-4.45	-3.14	0.47
Ser	UCA	<sup>d</sup> —	<b>1.76</b>	<sup>d</sup> —
Ser	UCG	<sup>d</sup> —	<sup>d</sup> —	<sup>d</sup> —
Ser	AGU	-4.79	<b>1.35</b>	-0.61
Ser	AGC	<b>1.24</b>	-0.25	-1.53
Pro	CCU	-0.48	-2.90	0.33
Pro	CCC	-2.85	-0.28	0.33
Pro	CCA	-0.32	-4.40	0.31
Pro	CCG	-0.23	-2.69	0.28
Thr	ACU	-0.08	0.58	-0.21
Thr	ACC	-0.57	0.53	-0.05
Thr	ACA	0.58	<b>1.28</b>	-5.24
Thr	ACG	-4.40	<sup>d</sup> —	0.48
Ala	GCU	0.55	-2.56	0.00
Ala	GCC	-0.07	0.10	-0.01
Ala	GCA	0.78	0.16	-0.61
Ala	GCG	0.76	1.15	-4.18
Tyr	UAU	0.95	-1.60	-0.39
Tyr	UAC	0.21	<b>1.37</b>	-2.22
His	CAU	-4.21	<b>1.75</b>	-4.18
His	CAC	<sup>d</sup> —	<b>1.75</b>	-3.81
Gln	CAA	0.62	-0.82	-0.15
Gln	CAG	-2.74	0.63	0.07
Asn	AAU	-4.47	-0.58	0.38
Asn	AAC	<sup>d</sup> —	0.18	0.25
Lys	AAA	-0.90	<sup>d</sup> —	0.39
Lys	AAG	<b>1.18</b>	<sup>d</sup> —	-0.69
Asp	GAU	<sup>d</sup> —	-3.17	0.48
Asp	GAC	<sup>d</sup> —	<sup>d</sup> —	0.48
Glu	GAA	0.93	<sup>d</sup> —	-0.28
Glu	GAG	-1.15	-5.28	0.41
Cys	UGU	<sup>d</sup> —	-4.75	0.48
Cys	UGC	-5.41	0.04	0.28
Arg	CGU	-1.82	0.28	0.18

TABLE 1: Continued.

Amino acid	Codon	<sup>a</sup> P value	<sup>b</sup> P value	<sup>c</sup> P value
Arg	CGC	<sup>d</sup> —	<b>1.52</b>	-1.03
Arg	CGA	0.83	-2.13	-0.22
Arg	CGG	-0.04	0.16	-0.04
Arg	AGA	<b>1.37</b>	-4.30	-1.32
Arg	AGG	<sup>d</sup> —	0.76	0.03
Gly	GGU	0.52	<sup>d</sup> —	0.04
Gly	GGC	-0.34	-4.61	0.32
Gly	GGA	-4.76	-4.55	0.48
Gly	GGG	<sup>d</sup> —	-4.62	0.48
Met	AUG	<b>1.01</b>	-0.35	-0.72
Trp	UGG	0.45	0.66	-0.61

<sup>a</sup>represents  $\alpha$ -helix.

<sup>b</sup>represents  $\beta$ -strand.

<sup>c</sup>represents The coil.

<sup>d</sup>represents The corresponding codon is not selected in the specific secondary structure unit.

Italic indicates that the corresponding codon has a weak bias to be selected in a specific secondary structure unit.

Bold indicates that the corresponding codon has a tendency to be selected in a specific secondary structure unit.

exist in the specific folding units of the nspl $\alpha$ . Based on the data of SCUB in the specific folding units (Table 1), the corresponding codon with  $N_1$  context was found to have a trend to exist in the specific folding unit of the nspl $\alpha$ . In detail, the codons with  $N_1$  context or  ${}_1N$  context (GUA~A, AGC~A, AAG~C, AGA~C, A~AUA, U~AGC, U~AAG, C~AAG, G~AGA, and U~AUG) have an obvious trend to exist in the helix unit of the nspl $\alpha$ . Some codons with  $N_1$  context or  ${}_1N$  context have a strong tendency to exist in the  $\beta$ -strand of the nspl $\alpha$ , including UUA~A, AUA~G, GUG~U, UCA~C, AGU~G, ACA~C, UAC~U, UAC~C, CAU~G, CAC~G, CGC~U, G~UUA, U~AUA, U~GUG, G~GUG, U~UCA, C~AGU, C~ACA, C~UAC, G~UAC, U~CAU, U~CAC, and U~CGC.

In order to identify the roles of nucleotide compositions (dinucleotide with  $N_1$  context and mononucleotide with  $N_1$  context) in shaping the codon with  $N_1$  context or  ${}_1N$  context, the  $R$  values for these interesting codons with  $N_1$  context or  ${}_1N$  context which have a strong tendency to exist in the helix or the  $\beta$ -strand were compared with the  $r$  values for the dinucleotide/mononucleotide with  $N_1$  context (Tables 3 and 4). The  $R$  value for the target codon with  $N_1$  context is higher than the corresponding dinucleotide/mononucleotide with  $N_1$  context or  ${}_1N$  context. For example, as for GUA which tends to exist in the helix of the nspl $\alpha$  gene, GUA~A has a tendency to exist in the helix unit, because the  $R$  value (1.4751) of GUA~A for the helix unit is higher than the  $R$  value for GUA~A for the  $\beta$ -strand and the coil (Table 2) and higher than the  $r$  value for UA~A and the  $r$  value for A~A (Tables 3 and 4). As for UUA which tends to exist in the  $\beta$ -strand of this gene, UUA~A has a tendency to exist in the strand unit, because the  $R$  value (4.9268) for UUA~A is higher than the  $R$  values for UUA~A of the helix and the coil and higher than the  $r$  values for UA~A and A~A (Tables 3 and 4). As for AGC which tends to exist in the helix of this gene, U~AGC has

TABLE 2: Relative abundance of codons with  $N_1$  context or  ${}_1N$  context in the PRRSV *nsplα* gene.

Codon-context ( $xyz-n$ )	R value <sup>1</sup>	R value <sup>2</sup>	R value <sup>3</sup>	Codon-context ( $xyz-n$ )	R value <sup>1</sup>	R value <sup>2</sup>	R value <sup>3</sup>
UUA~A	0.0000	4.9268	0.3345	A~UUA	0.0000	0.0401	3.6799
UUA~U	0.2911	0.0000	0.0000	U~UUA	0.1456	0.0617	0.2945
UUA~C	2.8120	0.0000	0.0000	C~UUA	2.9124	0.0000	0.0000
UUA~G	0.0000	0.0000	3.7713	G~UUA	0.0000	3.7019	0.0000
AUC~A	0.0000	0.0000	0.0000	A~AUC	0.0000	0.0000	2.4430
AUC~U	0.0000	0.0000	3.1392	U~AUC	0.0000	0.2233	0.0116
AUC~C	0.0000	0.0000	0.0000	C~AUC	0.0000	3.4849	0.0439
AUC~G	0.0000	3.7945	0.4601	G~AUC	0.0000	0.0000	1.5563
AUA~A	0.0000	0.0000	0.0000	A~AUA	0.0000	0.0000	0.0000
AUA~U	0.0000	0.0000	0.0000	U~AUA	4.3672	3.4158	2.2091
AUA~C	3.0128	0.0000	0.5557	C~AUA	0.0000	0.3703	0.0000
AUA~G	0.0000	3.7945	3.5998	G~AUA	0.0000	0.0000	1.5428
GUA~A	1.4751	0.0000	0.0934	A~GUA	2.9502	3.2845	0.0000
GUA~U	1.0918	3.7954	0.0822	U~GUA	0.0000	0.0000	0.0000
GUA~C	0.7532	0.0000	4.2386	C~GUA	0.7532	1.2342	0.0000
GUA~G	0.9274	0.0000	0.0000	G~GUA	0.9274	0.0000	4.1141
GUG~A	0.0510	0.0000	0.0000	A~GUG	1.4283	0.0000	0.3526
GUG~U	0.2014	2.7198	0.0000	U~GUG	0.0629	1.8132	0.0239
GUG~C	1.4413	0.1874	4.4454	C~GUG	0.0174	0.4872	0.0000
GUG~G	1.7318	0.8833	0.0000	G~GUG	2.7367	1.4824	3.7249
UCA~A	0.0000	0.0000	0.0000	A~UCA	0.0000	0.0000	0.0000
UCA~U	0.0000	0.0000	0.0000	U~UCA	0.0000	3.7954	0.0000
UCA~C	0.0000	3.7027	0.0000	C~UCA	0.0000	0.0000	0.0000
UCA~G	0.0000	0.0000	0.0000	G~UCA	0.0000	0.0000	0.0000
AGU~A	0.0000	0.0000	0.0000	A~AGU	0.0000	0.0000	0.0000
AGU~U	0.0000	0.0000	0.0000	U~AGU	2.4694	0.0000	0.0000
AGU~C	0.0473	0.0000	0.0000	C~AGU	1.2619	3.7027	4.4454
AGU~G	3.6513	3.7945	4.1141	G~AGU	0.0583	0.0000	0.0000
AGC~A	2.9502	0.0000	0.0000	A~AGC	0.0000	0.0000	0.0000
AGC~U	0.0000	0.0000	0.0000	U~AGC	5.9004	0.0000	0.0000
AGC~C	0.0000	0.0000	0.0000	C~AGC	0.0000	3.7027	0.0000
AGC~G	1.8548	3.7945	0.0000	G~AGC	0.0000	0.0000	0.0000
ACA~A	0.0000	0.0000	0.0000	A~ACA	0.0000	0.0000	0.0000
ACA~U	0.0000	0.0000	0.0000	U~ACA	4.3672	0.0000	3.4431
ACA~C	0.0000	3.7027	0.0383	C~ACA	0.0000	3.7027	0.1150
ACA~G	3.7096	0.0000	4.0786	G~ACA	0.0000	0.0000	0.0000
UAC~A	5.6595	0.0000	3.9513	A~UAC	0.0000	0.0502	0.0000
UAC~U	0.0000	1.4455	0.0000	U~UAC	1.2478	0.8812	1.4434
UAC~C	0.1230	1.4253	0.0000	C~UAC	2.1520	1.3876	2.6300
UAC~G	0.0000	0.8887	0.0646	G~UAC	0.0000	1.4529	0.0000
CAU~A	0.0000	0.0000	0.0000	A~CAU	3.9336	0.0000	0.0000
CAU~U	0.0000	0.0000	0.0000	U~CAU	0.0000	3.7043	0.0000
CAU~C	1.0043	0.0000	0.0000	C~CAU	0.0000	0.0889	4.4454
CAU~G	2.4730	3.7945	4.1141	G~CAU	1.2365	0.0000	0.0000
CAC~A	0.0000	0.0000	0.0000	A~CAC	0.0000	0.0000	0.0000
CAC~U	0.0000	0.2138	0.0000	U~CAC	4.3672	3.5816	0.0000
CAC~C	0.0000	0.0000	0.0000	C~CAC	0.0000	0.2086	0.0000
CAC~G	3.7096	3.5807	0.0000	G~CAC	0.0000	0.0000	0.0000
AAG~A	0.0000	0.0000	3.4986	A~AAG	0.0000	0.0000	3.3416

TABLE 2: Continued.

Codon-context ( <i>xyz-n</i> )	<i>R</i> value <sup>1</sup>	<i>R</i> value <sup>2</sup>	<i>R</i> value <sup>3</sup>	Codon-context ( <i>xyz-n</i> )	<i>R</i> value <sup>1</sup>	<i>R</i> value <sup>2</sup>	<i>R</i> value <sup>3</sup>
AAG~U	0.0546	0.0000	0.0000	U~AAG	1.4739	0.0000	0.0000
AAG~C	2.9751	0.0000	0.0000	C~AAG	1.9960	0.0000	0.0000
AAG~G	0.0000	0.0000	0.5286	G~AAG	0.0000	0.0000	0.6895
CGC~A	1.3112	0.0000	0.0000	A~CGC	2.8704	0.0000	0.0000
CGC~U	0.1213	3.7954	0.0000	U~CGC	0.1180	2.3721	0.0000
CGC~C	0.0837	0.0000	0.0000	C~CGC	1.3843	1.3885	0.0000
CGC~G	2.6791	0.0000	0.0000	G~CGC	0.1003	0.0000	0.0000
AGA~A	0.0000	0.0000	1.9453	A~AGA	0.0831	0.0000	2.0578
AGA~U	0.1230	0.0000	0.1188	U~AGA	0.6766	0.0000	0.0099
AGA~C	2.8855	0.0000	0.0249	C~AGA	0.1273	0.0000	2.1293
AGA~G	0.0522	0.0000	1.9591	G~AGA	2.9259	0.0000	0.0230
AUG~A	0.0000	0.0000	2.1721	A~AUG	0.0000	0.0000	3.9812
AUG~U	0.0000	0.0000	0.0000	U~AUG	4.1823	3.6955	0.0000
AUG~C	0.0000	0.0000	0.0000	C~AUG	0.1275	0.0974	0.0000
AUG~G	0.0000	3.7945	1.8881	G~AUG	0.0000	0.0000	0.0340

<sup>1</sup>represents The relative abundance of codons with  $N_1$  context in the helix unit of the PRRSV *nsplα*.

<sup>2</sup>represents The relative abundance of codons with  $N_1$  context in the  $\beta$ -strand unit of the PRRSV *nsplα*.

<sup>3</sup>represents The relative abundance of codons with  $N_1$  context in the coil unit of the PRRSV *nsplα*.

a tendency to exist in the helix rather than in the strand or the coil, because the *R* value (5.9004) for U~AGC is higher than the *R* value of U~AGC of the strand and the coil and higher than the *r* values for U~AG and U~A (Tables 3 and 4). As for UUA which tends to exist in the strand of this gene, G~UUA has a tendency to exist in the strand unit, because the *R* value (3.7019) for G~UUA of the strand unit is higher than the *R* values for G~UUA of the helix and the coil and higher than *r* values for G~UU and G~U (Tables 3 and 4). Based on the standard mentioned above, GUA~A, AGC~A, AAG~C, AGA~C, A~GUA, U~AGC, U~AAG, C~AAG, G~AGA, and U~AUG have a strong trend to exist in the helix of PRRSV *nsplα* gene and UUA~A, AUA~G, GUG~U, UCA~C, ACA~C, UAC~U, UAC~C, CAU~G, CGC~U, G~UUA, U~GUG, U~UCA, C~AGU, C~ACA, G~UAC, U~CAU, and U~CGC have a strong tendency to exist in the  $\beta$ -strand of the *nsplα* gene.

#### 4. Discussion

In this study, we have mapped the fluctuation of the overall tRNA abundance for each codon position along the PRRSV *nsplα* gene and estimated the correlation between the synonymous codon usage and different folding units of the *nsplα*. The performance of mapping the fluctuation of the overall tRNA abundance for each codon position along the target gene likely reflects the translation speed of ribosomes scanning caused by the tRNA abundance of the pigs to some degree, since the tRNA abundance plays an important role in the ribosome scanning along the target coding sequence [35, 36]. The previous report showed that the  $\alpha$ -helix is preferentially coded by translationally fast mRNA regions while the slow segments often encode  $\beta$ -strands and coil regions [37]. In the study, no linkage between the fluctuation of the overall tRNA abundance pairing to the codon positions along

the *nsplα* gene and the specific folding units might suggest that the process of translation fine-tunes is not performed by variation of translation speed for each codon position along the *nsplα*. The fine-tuning *in vivo* protein folding exists in the gene, and this regularity is largely believed to occur in a cotranslational process [38]. However, the PRRSV *nsplα* derives from the posttranslational processing of the *pplα* [10, 39]. The process of the cleavage of the *nsplα* from the *pplα* polyprotein of PRRSV performed by the posttranslation might be free from the fluctuation of tRNA abundance pairing to the each codon position along the *nsplα* gene. As for the ribosomes scanning the *nsplα* gene, there is no significant link between the fluctuation of the overall tRNA abundance and the specific folding units, and the translation elongation rate of this gene is not high. These results suggest that the low tRNA abundance controls the ribosomal traffic along the translated message to achieve the effective synthesized product of the PRRSV *pplα*. The low translational elongation at the translation beginning step directs the target gene to generate the corresponding protein effectively [40].

Turning to the role of the synonymous codon usage in the formation of the specific units of the *nsplα*, there is significant relationship between the synonymous codon usage bias and the specific folding units in the target protein. The synonymous codons assist messenger RNA to carry the information of the specific folding units, and a single codon or a contiguous nucleotide region plays roles in shaping the specific folding units [24, 25, 41, 42]. As for the PRRSV *nsplα*, there is no synonymous codon which tends to exist in coil unit. However, many synonymous codons exist in the  $\alpha$ -helix and  $\beta$ -strands regions of this gene, and no synonymous codon has a strong tendency to be selected by both the  $\alpha$ -helix and the  $\beta$ -strands in the PRRSV *nsplα* simultaneously. These results indicated that SCUB might play roles in shaping this

TABLE 3: Relative abundance of dinucleotides with  $N_1$  context or  ${}_1N$  context in the PRRSV *nspl $\alpha$*  gene.

Dinucleotides with $N_1$ context ( $xy\sim n$ )	$r$ value	Dinucleotides with ${}_1N$ context $n\sim xy$ )	$r$ value
UC~A	0.6130	A~AU	1.4456
UC~U	1.3115	U~AU	0.8552
UC~C	1.2805	C~AU	0.6485
UC~G	0.5669	G~AU	1.0047
UA~A	0.6177	A~GU	1.2963
UA~U	1.0269	U~GU	1.1651
UA~C	1.7402	C~GU	0.5072
UA~G	0.4351	G~GU	1.1064
UG~A	1.2060	A~UC	0.5880
UG~U	0.8195	U~UC	1.0330
UG~C	1.0921	C~UC	1.1230
UG~G	0.8019	G~UC	1.0286
CA~A	1.8155	A~AG	1.2415
CA~U	0.4556	U~AG	0.2825
CA~C	0.9184	C~AG	0.9946
CA~G	0.8963	G~AG	1.5135
GU~A	0.3165	A~AC	1.5521
GU~U	0.9679	U~AC	0.9913
GU~C	0.8485	C~AC	0.8942
GU~G	1.5896	G~AC	0.6067
GC~A	1.2887	A~UA	0.6720
GC~U	1.0849	U~UA	0.9776
GC~C	1.3411	C~UA	1.6741
GC~G	0.3298	G~UA	0.5155
CA~A	1.8155	A~CA	0.9721
CA~U	0.4556	U~CA	0.4819
CA~C	0.9184	C~CA	1.3762
CA~G	0.8963	G~CA	1.0994
AC~A	0.9465	A~AA	1.1734
AC~U	1.2243	U~AA	0.3547
AC~C	0.8848	C~AA	1.7821
AC~G	0.9524	G~AA	1.4708
AU~A	0.5596	A~CG	1.6312
AU~U	0.9124	U~CG	0.7431
AU~C	0.6579	C~CG	1.2830
AU~G	1.7819	G~CG	0.4693
AG~A	1.0440	A~UU	0.8843
AG~U	1.3708	U~UU	1.5257
AG~C	0.7761	C~UU	1.2791
AG~G	0.8164	G~UU	1.2724
GC~A	1.2887		
GC~U	1.0849		
GC~C	1.3411		
GC~G	0.3298		

TABLE 4: Relative abundance of mononucleotide with  $N_1$  context in PRRSV *nspl $\alpha$*  gene.

Mononucleotide with $N_1$ context	$r(x\sim n)$
A~A	1.0664
A~U	0.7354
A~C	1.0751
A~G	0.9433
U~A	0.6124
U~U	0.7588
U~C	0.8228
U~G	1.4182
C~A	1.0468
C~U	1.1225
C~C	0.8788
C~G	0.6277
G~A	1.0557
G~U	0.9975
G~C	0.8930
G~G	0.7782

protease with natural properties for the life-cycle of PRRSV. SCUB for formation of the specific folding units of the PRRSV *nspl $\alpha$*  is influenced by the natural selection. As an example of the role for natural selection, the expressivity of genes is an important factor in shaping SCUB, both for prokaryotic and for eukaryotic organisms [18, 22, 43, 44]. Although the link between the SCUB and the formation of the specific folding units was reported [25, 35, 37, 38, 45], the role of CDCB in formation of specific folding units is not clear. In this study, we found that CDCB plays a role in the formation of specific folding units in the PRRSV *nspl $\alpha$* . The synonymous usage bias and CDCB, which play important roles in achieving accuracy and efficiency in protein synthesis, are particular manifestations of coding sequence nonrandomness [23, 46, 47]. Spatial interaction of ribosomal proteins with codon-anticodon RNA pairs inside the A and P sites of the ribosome could be preferable for particular codons with context [20, 48].

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Authors' Contribution

Yao-zhong Ding, Ya-nan You, and Dong-jie Sun contributed to the original draft of the paper and approved the final version. Hao-tai Chen, Yong-lu Wang, Hui-yun Chang, Li Pan, Yu-zhen Fang, Zhong-wang Zhang, Peng Zhou, Jian-liang Lv, Xinsheng Liu, Jun-jun Shao, Fu-rong Zhao, and Tong Lin downloaded the sequences and analyzed the data. Laszlo Stipkovits, Zygmunt Pejsak, Yong-guang Zhang, and Jie Zhang provided suggestive information to the Discussion

and revised the paper. All authors read and approved the final version.

## Acknowledgments

This work was supported in part by Grants from International Science and Technology Cooperation Program of China (no. 2012DFG31890) and Gansu Provincial Funds for Distinguished Young Scientists (1111RJDA005). This study was also supported by National Natural Science foundation of China (no. 31172335 and no. 31072143).

## References

- [1] J. G. Cho and S. A. Dee, "Porcine reproductive and respiratory syndrome virus," *Theriogenology*, vol. 66, no. 3, pp. 655–662, 2006.
- [2] K.-K. Conzelmann, N. Visser, P. van Woensel, and H.-J. Thiel, "Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group," *Virology*, vol. 193, no. 1, pp. 329–339, 1993.
- [3] J. J. M. Meulenbergh, M. M. Hulst, E. J. De Meijer et al., "Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV," *Virology*, vol. 192, no. 1, pp. 62–72, 1993.
- [4] R. Allende, T. L. Lewis, Z. Lu et al., "North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions," *Journal of General Virology*, vol. 80, part 2, pp. 307–315, 1999.
- [5] E. M. Bautista, S. M. Goyal, I. J. Yoon, H. S. Joo, and J. E. Collins, "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*, vol. 5, no. 2, pp. 163–165, 1993.
- [6] J. E. Collins, D. A. Benfield, W. T. Christianson 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," *Journal of Veterinary Diagnostic Investigation*, vol. 4, no. 2, pp. 117–126, 1992.
- [7] Y. S. Liu, J. H. Zhou, H. T. Chen et al., "Analysis of synonymous codon usage in porcine reproductive and respiratory syndrome virus," *Infection, Genetics and Evolution*, vol. 10, no. 6, pp. 797–803, 2010.
- [8] C. J. Nelsen, M. P. Murtaugh, and K. S. Faaberg, "Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents," *Journal of Virology*, vol. 73, no. 1, pp. 270–280, 1999.
- [9] E. J. Snijder, A. L. M. Wassenaar, L. C. Van Dinten, W. J. M. Spaan, and A. E. Gorbalenya, "The arterivirus Nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases," *Journal of Biological Chemistry*, vol. 271, no. 9, pp. 4864–4871, 1996.
- [10] Y. Fang and E. J. Snijder, "The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins," *Virus Research*, vol. 154, no. 1-2, pp. 61–76, 2010.
- [11] D. D. Nedialkova, A. E. Gorbalenya, and E. J. Snijder, "Arterivirus Nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs," *PLOS Pathogens*, vol. 6, no. 2, Article ID e1000772, 2010.
- [12] J. A. den Boon, K. S. Faaberg, J. J. M. Meulenbergh et al., "Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases," *Journal of Virology*, vol. 69, no. 7, pp. 4500–4505, 1995.
- [13] E. J. Snijder, A. L. M. Wassenaar, and W. J. M. Spaan, "The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease," *Journal of Virology*, vol. 66, no. 12, pp. 7040–7048, 1992.
- [14] T. Dokland, "The structural biology of PRRSV," *Virus Research*, vol. 154, no. 1-2, pp. 86–97, 2010.
- [15] M. V. Kroese, J. C. Zevenhoven-Dobbe, J. N. A. Bos-de Ruijter et al., "The nsp 1 $\alpha$  and nsp 1 $\beta$  papain-like autoproteases are essential for porcine reproductive and respiratory syndrome virus RNA synthesis," *Journal of General Virology*, vol. 89, no. 2, pp. 494–499, 2008.
- [16] S. Yuna, X. Fei, G. Yu et al., "Crystal structure of porcine reproductive and respiratory syndrome virus leader protease Nsp1 $\alpha$ ," *Journal of Virology*, vol. 83, no. 21, pp. 10931–10940, 2009.
- [17] X. Shi, J. Chen, G. Xing et al., "Amino acid at position 176 was essential for porcine reproductive and respiratory syndrome virus (PRRSV) non-structural protein 1 $\alpha$  (nsp1 $\alpha$ ) as an inhibitor to the induction of IFN- $\beta$ ," *Cellular Immunology*, vol. 280, no. 2, pp. 125–131, 2012.
- [18] I. Bahir, M. Fromer, Y. Prat, and M. Linial, "Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences," *Molecular Systems Biology*, vol. 5, article 311, 2009.
- [19] T. F. Clarke IV and P. L. Clark, "Rare codons cluster," *PLoS ONE*, vol. 3, no. 10, Article ID e3412, 2008.
- [20] A. Fedorov, S. Saxonov, and W. Gilbert, "Regularities of context-dependent codon bias in eukaryotic genes," *Nucleic Acids Research*, vol. 30, no. 5, pp. 1192–1197, 2002.
- [21] J. Zhou, Z. Gao, J. Zhang et al., "The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts," *Infection, Genetics and Evolution*, vol. 14, no. 1, pp. 105–110, 2013.
- [22] J. Zhou, J. Su, H. Chen et al., "Clustering of low usage codons in the translation initiation region of hepatitis C virus," *Infection, Genetics and Evolution*, vol. 18, pp. 8–12, 2013.
- [23] H. Akashi, "Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy," *Genetics*, vol. 136, no. 3, pp. 927–935, 1994.
- [24] I. Weygand-Durasevic and M. Ibba, "Cell biology: new roles for codon usage," *Science*, vol. 329, no. 5998, pp. 1473–1474, 2010.
- [25] J. h. Zhou, Y. N. You, H. T. Chen et al., "The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus," *Infection, Genetics and Evolution*, vol. 16, pp. 270–274, 2013.
- [26] T. Zhou, M. Weems, and C. O. Wilke, "Translationally optimal codons associate with structurally sensitive sites in proteins," *Molecular Biology and Evolution*, vol. 26, no. 7, pp. 1571–1580, 2009.
- [27] Y. Guisez, J. Robbens, E. Remaut, and W. Fiers, "Folding of the MS2 coat protein in *Escherichia coli* is modulated by translational pauses resulting from mRNA secondary structure and codon usage: a hypothesis," *Journal of Theoretical Biology*, vol. 162, no. 2, pp. 243–252, 1993.
- [28] X. Tao and D. Dafu, "The relationship between synonymous codon usage and protein structure," *FEBS Letters*, vol. 434, no. 1-2, pp. 93–96, 1998.
- [29] M. Gouy, "Codon contexts in enterobacterial and coliphage genes," *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 426–444, 1987.

- [30] S. Karlin and J. Mrázek, "What drives codon choices in human genes?" *Journal of Molecular Biology*, vol. 262, no. 4, pp. 459–472, 1996.
- [31] E. G. Shpaer, "Constraints on codon context in *Escherichia coli* genes. Their possible role in modulating the efficiency of translation," *Journal of Molecular Biology*, vol. 188, no. 4, pp. 555–564, 1986.
- [32] O. G. Berg and P. J. N. Silva, "Codon bias in *Escherichia coli*: The influence of codon context on mutation and selection," *Nucleic Acids Research*, vol. 25, no. 7, pp. 1397–1404, 1997.
- [33] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.
- [34] S. Karlin and C. Burge, "Dinucleotide relative abundance extremes: a genomic signature," *Trends in Genetics*, vol. 11, no. 7, pp. 283–290, 1995.
- [35] J. L. Parmley and M. A. Huynen, "Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation," *PLoS Genetics*, vol. 5, no. 7, Article ID e1000548, 2009.
- [36] E. P. C. Rocha, "Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization," *Genome Research*, vol. 14, no. 11, pp. 2279–2286, 2004.
- [37] T. A. Thanaraj and P. Argos, "Protein secondary structural types are differentially coded on messenger RNA," *Protein Science*, vol. 5, no. 10, pp. 1973–1983, 1996.
- [38] A. A. Komar, "A pause for thought along the co-translational folding pathway," *Trends in Biochemical Sciences*, vol. 34, no. 1, pp. 16–24, 2009.
- [39] D. van Aken, J. Zevenhoven-Dobbe, A. E. Gorbalenya, and E. J. Snijder, "Proteolytic maturation of replicase polyprotein ppla by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7," *Journal of General Virology*, vol. 87, no. 12, pp. 3473–3482, 2006.
- [40] T. Tuller, A. Carmi, K. Vestsigian et al., "An evolutionarily conserved mechanism for controlling the efficiency of protein translation," *Cell*, vol. 141, no. 2, pp. 344–354, 2010.
- [41] D. B. Carlini, Y. Chen, and W. Stephan, "The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh* and *Adhr*," *Genetics*, vol. 159, no. 2, pp. 623–633, 2001.
- [42] R. Saunders and C. M. Deane, "Synonymous codon usage influences the local protein structure observed," *Nucleic Acids Research*, vol. 38, no. 19, pp. 6719–6728, 2010.
- [43] M. D. Ermolaeva, "Synonymous codon usage in bacteria," *Current Issues in Molecular Biology*, vol. 3, no. 4, pp. 91–97, 2001.
- [44] J. B. Plotkin, H. Robins, and A. J. Levine, "Tissue-specific codon usage and the expression of human genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 34, pp. 12588–12591, 2004.
- [45] G. Zhang and Z. Ignatova, "Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis," *PLoS ONE*, vol. 4, no. 4, Article ID e5036, 2009.
- [46] A. Eyre-Walker, "Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy?" *Molecular Biology and Evolution*, vol. 13, no. 6, pp. 864–872, 1996.
- [47] J. Precup and J. Parker, "Missense misreading of asparagine codons as a function of codon identity and context," *The Journal of Biological Chemistry*, vol. 262, no. 23, pp. 11351–11355, 1987.
- [48] R. Green and H. F. Noller, "Ribosomes and translation," *Annual Review of Biochemistry*, vol. 66, pp. 679–716, 1997.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

