



The effects of codon usage on the formation of secondary structures of nucleocapsid protein of peste des petits ruminants virus

Xiao-xia Ma^{1,2,3} · Yi-ning Wang¹ · Xiao-an Cao¹ · Xue-rui Li¹ · Yong-sheng Liu¹ · Jian-hua Zhou¹ · Xue-peng Cai¹

Received: 30 January 2018 / Accepted: 23 March 2018 / Published online: 31 May 2018
© The Genetics Society of Korea and Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

The nucleocapsid (N) protein of peste des petits ruminants virus (PPRV) with a conserved amino acid usage pattern plays an important role in viral replication. The primary objective of this study was to estimate roles of synonymous codon usages of PPRV N gene and tRNA abundances of host in the formation of secondary structure of N protein. The potential effects of synonymous codon usages of N gene and tRNA abundances of host on shaping different folding units (α -helix, β -strand and the coil) in N protein were estimated, based on the information about the modeling secondary structure of PPRV N protein. The synonymous codon usage bias was found in different folding units in PPRV N protein. To better understand the role of translation speed caused by variant tRNA abundances in shaping the specific folding unit in N protein, we modeled the changing trends of tRNA abundance at the transition boundaries from one folding unit to another folding unit (β -strand \rightarrow coil, coil \rightarrow β -strand, α -helix \rightarrow coil, coil \rightarrow α -helix). The obvious fluctuations of tRNA abundance were identified at the two transition boundaries (β -strand \rightarrow coil and coil \rightarrow β -strand) in PPRV N protein. Our findings suggested that viral synonymous codon usage bias and cellular tRNA abundance variation might have potential effects on the formation of secondary structure of PPRV N protein.

Keywords Peste des petits ruminants virus · PPRV · Nucleocapsid · Synonymous codon usage · tRNA abundance · Secondary structure

Xiao-xia Ma and Yi-ning Wang have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13258-018-0684-2>) contains supplementary material, which is available to authorized users.

- ✉ Xiao-xia Ma
maxiaoxia956@163.com
- ✉ Jian-hua Zhou
zhoujianhua03@caas.cn
- ✉ Xue-peng Cai
caixuepeng@caas.cn

- ¹ State Key Laboratory of Veterinary Etiological Biology, National Foot-and-Mouth Disease Reference Laboratory, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, Gansu, People's Republic of China
- ² Engineering & Technology Research Center for Animal Cell, Northwest Minzu University, Lanzhou, Gansu, People's Republic of China
- ³ College of Life Science and Engineering, Northwest Minzu University, Lanzhou, Gansu, People's Republic of China

Introduction

Peste des petits ruminants virus (PPRV), which was classified as a *Morbillivirus* in the family *Paramyxoviridae*, is considered as an important pathogen that contributes to a highly contagious disease in goats and sheep (Banyard et al. 2010; Gibbs et al. 1979). The PPRV is an enveloped single-stranded negative-sense RNA virus (Chard et al. 2008). There were four distinct lineages of PPRV, namely lineage I, II, III and IV, based on the geographical distribution of PPRV, but only one serotype of PPRV was found (Banyard et al. 2010). The PPRV genome consists of six transcriptional units (N, P, M, F, H and L genes) which can synthesize the nucleoprotein (N), polymerase complex (P), matrix protein (M), fusion glycoprotein (F), haemagglutinin glycoprotein (H) and large polymerase (L) (Bailey et al. 2007). According to the schematic diagram of PPRV virion structure (Banyard et al. 2010), viral genome was closely encapsulated by N protein to constitute a helical nucleocapsid, like other negative stranded RNA viruses. As for PPRV N protein, the molecular weight was about 58 kDa (Ismail

et al. 1995). This protein was considered as a major viral protein in *Morbilliviruses* (Diallo et al. 1987), and it played an key role in PPRV replication (Parida et al. 2015). Interestingly, PPRV N gene can be expressed in insect, bacterial and mammalian cells and the product of the gene expression contributes to formation of nucleocapsid-like particle (Mitra-Kaushik et al. 2001). Meanwhile, the conserved region of N protein functioned as the important requirement for self-assembly of measles virus (a member of *Morbillivirus*) (Bankamp et al. 1996). Like N protein of measles virus, the conserved nucleotide sequence of N protein was a key factor in the formation of virus-like particles of PPRV (Liu et al. 2014). The findings implied that the specific nucleotide usage pattern of N gene would function as the formation of secondary structure of N protein. Currently, the analysis of viral protein structure has been an important pathway to find out biological and immune functions and improve vaccine designs. Unfortunately, there was no report about any viral proteins of PPRV. This current situation might have a limit in estimating the potential effects of genetic information from gene sequence on the formation of viral protein structure of PPRV and lead to a lack of relationship between gene sequence and protein secondary structure in vaccine design. Although the genetic features of PPRV N gene were significantly different from those of other viral genes (H, F, L, P and M genes), the N genes of *Morbillivirus* represented highly conserved genetic features (Baron et al. 2016). Moreover, the studies for secondary structure of N protein of measles virus have been carried out, and the related precise information about secondary structures of N protein can benefit for investigations into structural features and biological functions of N proteins of other members of *Morbillivirus*. The nucleotide usage pattern had obvious effects on synonymous codon usage patterns of PPRV genes, and synonymous codon usage patterns of N gene distinguished from that of other viral genes (Ma et al. 2017). The synonymous codon usage pattern was identified as an important genetic factor in impact on evolutionary dynamic, translation efficiency, tRNA abundance and secondary structure of protein, etc. (Bahir et al. 2009; Clarke and Clark 2008; dos Reis et al. 2003; Singh et al. 2017; Straube 2017; Welch et al. 2009; Weygand-Durasevic and Ibba 2010). It was found that synonymous codon usage bias profoundly influenced the formation of secondary structures to regulate biological function of native protein (Chaney and Clark 2015; Rodriguez et al. 2017). Some previous studies pointed out that various tRNA abundances could reflect synonymous codon usage patterns in the cell environment (Hanson and Coller 2017; Mioduser et al. 2017; Nossmann et al. 2017; Rafels-Ybern et al. 2017). Subsequently, some reports represented that the synonymous codon usage bias had obvious correlations with the tRNA abundance in host cells (Kanduc 2017; Pouyet et al. 2017). Compared with the

obvious effect of amino acid usage patterns on the formation of secondary structure of native protein, synonymous codon usage bias had a potential to regulate the correct folding units in protein secondary structure (Gu et al. 2004; Guisez et al. 1993; Marin 2008). Both theoretical in silico modeling analyses and the experimental data indicated that nucleotide and codon usages at mRNA levels had additional layers of information on fine-tunes in vivo protein folding during co-translational process, beyond the amino acid usages (Komar 2009). Analyses of synonymous codon usage bias in different folding units (α -helix, β -strand and the coil) and changes of translation speed at the transition boundary from one folding unit to another folding unit might give a new insight into the formation of viral protein product in host cells. According to the effect of synonymous codon usage bias on protein secondary structure, the study for synonymous codon usage bias in different folding units of native protein would benefit for providing new insights into biological functions or vaccine design related to PPRV N protein.

Materials and methods

The coding sequences of PPRV N protein

The 45 genomes PPRV were downloaded from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>) and the detailed information was listed in Table S1. To better identify unique genetic feature reflected by the overall codon usage bias between N gene and other viral genes (H, F, L, M and P genes), an effective number of codons (ENC) analysis was applied to quantify the absolute codon usage bias for these PPRV genes (Wright 1990). Meanwhile, to better clarify the adaptation of PPRV N gene to host, an improved CAI calculation method, depending on CAIcal server, was applied for analyzing extents of adaptation of viral genes (N, H, F, L, M and P genes) to host (*Ovis aries*) (Puigbo et al. 2008a; Sharp and Li 1987). The codon usage frequencies of *O. aries* (natural host of PPRV) was selected for a reference and the related data was obtained from the Codon Usage Database (Nakamura et al. 2000). To better clarify genetic features reflected by the overall codon usage bias and its adaptation to host for each viral gene, the ENC values and CAI values of the six genes were estimated by the statistical test (One-Way ANOVA) via SPSS 16.0 software for Windows, respectively.

Alignment of amino acid sequence of PPRV N protein and the model of N protein structure

To identify genetic diversity of amino acid usages in PPRV N protein, the genetic diversity of N genes in 45 PPRV strains was analyzed by MEGA 5.0 software. Phylogenetic

tree of N protein was constructed using the Neighbor-Joining method via gamma distributed rate variation and a boot-strap of 1000 replicates.

There were two reliable biological conclusions: (i) the N proteins represented highly conserved genetic features in *Morbillivirus*, (ii) the precise secondary structure of N protein in measles virus was obtained by X-ray diffraction method (Guryanov et al. 2015; Gutsche et al. 2015). The information benefited for modeling the reliable secondary structure of PPRV N protein. In this study, each PPRV N gene was estimated and modeled for its specific 3D structure via SWISS-MODEL program which is supported by Swiss Institute of Bioinformatics and Center for Molecular Life Sciences (<https://swissmodel.expasy.org/>). SWISS-MODEL program relied on a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb Viewer).

Modeling the change of tRNA abundance at the transmission boundary from one specific folding unit to another one

Based on the information on the 3D structure of PPRV N protein deriving from SWISS-MODEL program, the specific folding unit can be located in amino acid sequence of each N protein. According to the previous reports regarding to the effect of synonymous codon usage bias on the formation of the specific folding unit in the target protein (Ding et al. 2014; Zhou et al. 2013b), the correlation between synonymous codon usage bias and the specific folding unit in N protein can be calculated by the formula below:

$$P = \frac{f_{obs}}{f_{exp}}$$

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

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

where $N_{(i,sec-k)}$ represents the amount of a particular synonymous codon coding for the corresponding amino acid in a specific secondary unit of protein, $sec-k$ represents the corresponding amino acid in the interesting secondary unit (the α -helix, β -strand or coil); $N_{(k)}$ represents the amount of the corresponding amino acid in the interesting secondary unit. In addition, $\sum N_{(i,sec-j)}$ represents the total number of amino acid in the specific secondary unit, $sec-j$ corresponds to a certain type of the three secondary structure units (α -helix, β -strand or coil), and N_{total} represents the total number of

codon in the target protein. Furthermore, we defined that when P value was much higher than 1.5, the synonymous codon had a strong tendency to exist in the interesting secondary unit; on the contrary, when P value was much less than 0.5, the synonymous codon had a strong tendency to avoid the interesting secondary unit (Zhou et al. 2013b).

Mapping changing trends of tRNA abundance at the transition boundary

To better model the translation speed caused by tRNA abundances at the transition boundary from one folding unit to another, we depended on the relative synonymous codon usage value (RSCU) of *O. aries* (natural host of PPRV) (Zhou et al. 2013a) in order to reflect the corresponding tRNA abundance, and therefore carried out for the formula below:

$$R = \sqrt[n]{\prod_1^n (W_{ij}/W_j)}$$

where R value represents that the overall codon usage bias for a particular codon position in the interesting gene, W_{ij} represents the specific synonymous codon (i) for the corresponding amino acid (j), W_j represents the highest RSCU of a synonymous codon for the same amino acid and n represents means of the given PPRV N coding sequences. A codon site with R value close to 1.00 is made by high tRNA abundance, while one with R value close to 0.00 is made by low tRNA abundance.

Results

Unique genetic feature of codon and amino acid usages for N gene

To estimate the magnitude of the codon usage bias of viral genes, the ENC values were calculated for all PPRV strains. Although the overall codon usage bias of N gene was significant different from those of F, H, L and P genes, the ENC data of the six genes of PPRV were more than 50 (Fig. 1a). The result suggested that the relatively weak synonymous codon usage bias of PPRV genes was strongly influenced by mutation pressure from viral nucleotide composition. Furthermore, to better quantify the magnitude of the adaptation of viral genes to host represented by CAI data, this magnitude (0.66 ± 0.007) of N gene was highest and had significant difference from those of the reset genes (p value < 0.001) (Fig. 1b). Because CAI analysis was able to reflect the adaptation of viral genes to the host cellular machinery (Butt et al. 2016), the result implied that PPRV N gene had a better

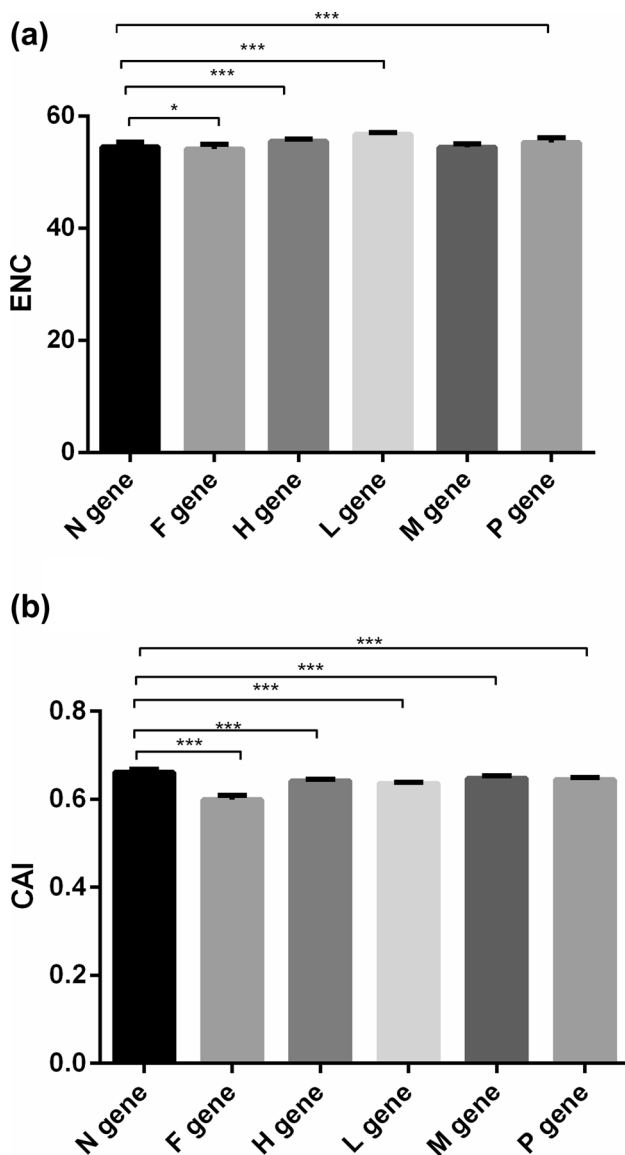


Fig. 1 The codon usage bias for the six genes of PPRV analyzed by ENC and CAI methods. **a** ENC data of the six genes, **b** CAI data of the six genes. CAI analysis of viral genes in relation to its host (*O. aries*). The p value was calculated by One-way ANOVA method of SPSS software. When p value < 0.05, the two groups have significant differences. ***means p value < 0.001, **means p value < 0.01, *means p value < 0.05

fitness to host and higher gene expression than other viral genes.

Turning to amino acid usages of N protein among different PPRV strains, the sequences were similar to each other (more than 95% similarity in amino acid sequence) (Fig. S1) and did not contain any insertion or deletion in the given sequences, suggesting that PPRV N protein was highly conserved and can be modeled for its 3D structure. According to the modeling 3D structure of PPRV N protein modeled by SWISS-MODEL program (Fig. S2), PPRV N protein

was rich in the α -helix units (AA₄₆–AA₅₉, AA₆₆–AA₇₆, AA₈₄–AA₉₁, AA₁₂₄–AA₁₃₀, AA₁₆₀–AA₁₈₂, AA₁₉₀–AA₂₀₀, AA₂₁₃–AA₂₂₃, AA₂₂₇–AA₂₄₂, AA₂₄₉–AA₂₄₆, AA₂₆₈–AA₂₇₈, AA₂₉₁–AA₃₀₅, AA₃₁₉–AA₃₂₄, AA₃₃₀–AA₃₄₄, AA₃₆₀–AA₃₈₃, AA₃₈₈–AA₄₀₃), of which the specific structure was capable to benefit for viral RNA sequence binding to this protein.

Codon usage bias in different folding units of PPRV N protein

Based on P values for the 61 codons which were involved in the formation of different folding units in N protein, we estimated the effect of synonymous codon usage pattern on the specific folding unit. In detail, some synonymous codons had propensity to exist in the α -helix unit, including AUU for Ile, GUC for Val, UCU/AGC for Ser, UAU for Tyr, UGU/UGC for Cys, CGG for Arg, AUG for Met and UGG for Trp, however, there were some synonymous codons which tended to avoid the formation of this folding unit, namely CCU, CCA/CCG for Pro, ACA/ACG for Thr, GAU for Asp, CGC for Arg and GGC for Gly (Table 1). As for codons usage with a bias for shaping the β -strand unit, these codons tended to be selected, namely CUU for Leu, AUU/AUA for Ile, GUU/GUC/GUA for Val, AGU for Ser, CCC/CCG for Pro, ACC for Thr, AAU for Asn, CGC for Arg, GGG for Gly, UGG for Trp, while other codons had strong tendency to avoid in this unit, including UUC for Phe, CUC/CUA for Leu, GUG for Val, UCU/UCC/UCA/UCG/AGC for Ser, CCU/CCA for Pro, ACU/ACA/ACG for Thr, GCU/GCC/GCG for Ala, UAU/UAC for Tyr, CAU/CAC for His, CAG for Gln, GAU for Asp, GAG for Glu, UGU/UGC for Cys, CGU/CGA/CGG/AGG for Arg (Table 1). Turning to some synonymous codons usage with link to the formation of the coil unit, there were CCU/CCA for Pro, ACA/ACG for Thr, CAU/CAC for His, GAU for Asp, GGC for Gly strongly selected by this unit, but the codons (CUU for Leu, AUU for Ile, GUA/GUC for Val, UGU/UGC for Cys, CGC for Arg, UGG for Trp) were slightly selected by this unit (Table 1). Compared with different P values for the same codon which was selected by different folding units, AUU for Ile, GUC for Val and UGG for Trp had an obvious tendency to shape the α -helix and β -strand, but the same codons were not able to form the three different units at the same time (Table 1). Of note, the two synonymous codons (UGU/UGC) coding for Cys had a significant relation with the formation of the α -helix; the two codons (CGC for Arg and UGG for Trp) had the obvious tendency to shape the β -strand unit; the two synonymous codons (CAU/CAC coding for His) had a strong tendency for the formation of the coil unit. The result represented that the selection of some synonymous codons was particularly involved in the formation of the different secondary units in the PPRV N protein to some degree.

Table 1 The effect of codon usage pattern on the different secondary structure units in PPRV N protein

Amino acid	Codon	<i>P</i> value for α -helix	<i>P</i> value for β -strand	<i>P</i> value for coil
Phe	UUU	0.782	1.369	1.103
Phe	UUC	1.106	<u>0.000</u>	1.066
Leu	UUA	1.211	0.066	0.982
Leu	UUG	1.431	1.182	0.668
Leu	CUU	1.339	3.168	<u>0.453</u>
Leu	CUC	1.079	<u>0.000</u>	1.085
Leu	CUA	0.976	<u>0.223</u>	1.127
Leu	CUG	1.340	1.186	0.732
Ile	AUU	1.532	2.112	<u>0.465</u>
Ile	AUC	1.245	1.463	0.760
Ile	AUA	1.129	1.787	0.797
Val	GUU	0.841	4.541	0.613
Val	GUC	1.613	2.253	<u>0.387</u>
Val	GUA	1.385	2.920	<u>0.455</u>
Val	GUG	1.191	<u>0.423</u>	0.946
Ser	UCU	1.699	<u>0.000</u>	0.645
Ser	UCC	0.879	<u>0.000</u>	1.227
Ser	UCA	1.067	<u>0.000</u>	1.094
Ser	UCG	1.240	<u>0.000</u>	0.971
Ser	AGU	0.878	4.319	0.617
Ser	AGC	1.786	<u>0.133</u>	0.564
Pro	CCU	<u>0.045</u>	<u>0.113</u>	1.804
Pro	CCC	1.129	2.013	0.765
Pro	CCA	<u>0.000</u>	<u>0.089</u>	1.840
Pro	CCG	<u>0.051</u>	3.535	1.316
Thr	ACU	1.042	<u>0.000</u>	1.111
Thr	ACC	0.664	1.581	1.157
Thr	ACA	<u>0.467</u>	<u>0.000</u>	1.520
Thr	ACG	<u>0.000</u>	<u>0.000</u>	1.852
Ala	GCU	1.000	<u>0.000</u>	1.141
Ala	GCC	1.156	<u>0.000</u>	1.031
Ala	GCA	1.162	1.384	0.831
Ala	GCG	0.755	<u>0.167</u>	1.292
Tyr	UAU	1.620	<u>0.000</u>	0.700
Tyr	UAC	0.875	<u>0.000</u>	1.230
His	CAU	<u>0.000</u>	<u>0.000</u>	1.852
His	CAC	<u>0.000</u>	<u>0.000</u>	1.852
Gln	CAA	0.751	1.241	1.143
Gln	CAG	0.995	<u>0.000</u>	1.145
Asn	AAU	1.022	4.197	0.533
Asn	AAC	0.612	0.764	1.309
Lys	AAA	0.790	0.550	1.213
Lys	AAG	1.116	0.650	0.967
Asp	GAU	<u>0.061</u>	<u>0.168</u>	1.785
Asp	GAC	1.040	1.267	0.934
Glu	GAA	1.254	<u>0.000</u>	0.961
Glu	GAG	0.903	1.033	1.064
Cys	UGU	2.084	<u>0.000</u>	<u>0.370</u>

Table 1 (continued)

Amino acid	Codon	<i>P</i> value for α -helix	<i>P</i> value for β -strand	<i>P</i> value for coil
Cys	UGC	2.606	<u>0.000</u>	<u>0.000</u>
Arg	CGU	1.288	<u>0.072</u>	0.926
Arg	CGC	<u>0.000</u>	9.792	<u>0.468</u>
Arg	CGA	0.709	<u>0.000</u>	1.348
Arg	CGG	1.811	<u>0.093</u>	0.552
Arg	AGA	0.949	0.803	1.064
Arg	AGG	0.867	<u>0.192</u>	1.209
Gly	GGU	0.859	<u>0.000</u>	1.242
Gly	GGC	<u>0.172</u>	<u>0.000</u>	1.730
Gly	GGA	0.586	0.803	1.322
Gly	GGG	0.809	1.636	1.046
Met	AUG	1.556	<u>0.000</u>	0.746
Trp	UGG	1.563	5.242	<u>0.000</u>

Underline 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

The fluctuations of translation speed caused by tRNA abundances from one type of folding unit to another

According to the information about 3D structure of N protein (Fig. S2), there were four types of transition boundaries from one folding unit to another folding unit, including between the coil unit and the β -strand unit and between the coil unit and the α -helix unit. To better model changing trends of translation speed caused by variant tRNA abundances, we depended on 59 synonymous codon usage values plus the two codons (AUG for Met and UGG for Trp) usage values being 1.0 to reflect tRNA abundances corresponding to 61 codons, and therefore modeled changing trends of translation speed at the four types of transition boundary (Fig. 2). Compared with changing trends of translation speed at the two transition boundaries (between the coil unit and the α -helix unit) (Fig. 2c, d), the significant fluctuations occurred at the two transition boundaries (between the coil unit and the β -strand unit) (Fig. 2a, b). The results suggested that the usages of synonymous codon might be affected by tRNA abundances, and the different translation speed caused by various tRNA abundances likely took part in the formation of the specific downstream folding unit.

Discussion

In this study, we evaluated the effects of synonymous codon usage on the formation of different folding units of PPRV N protein. There was a significant correlation between the

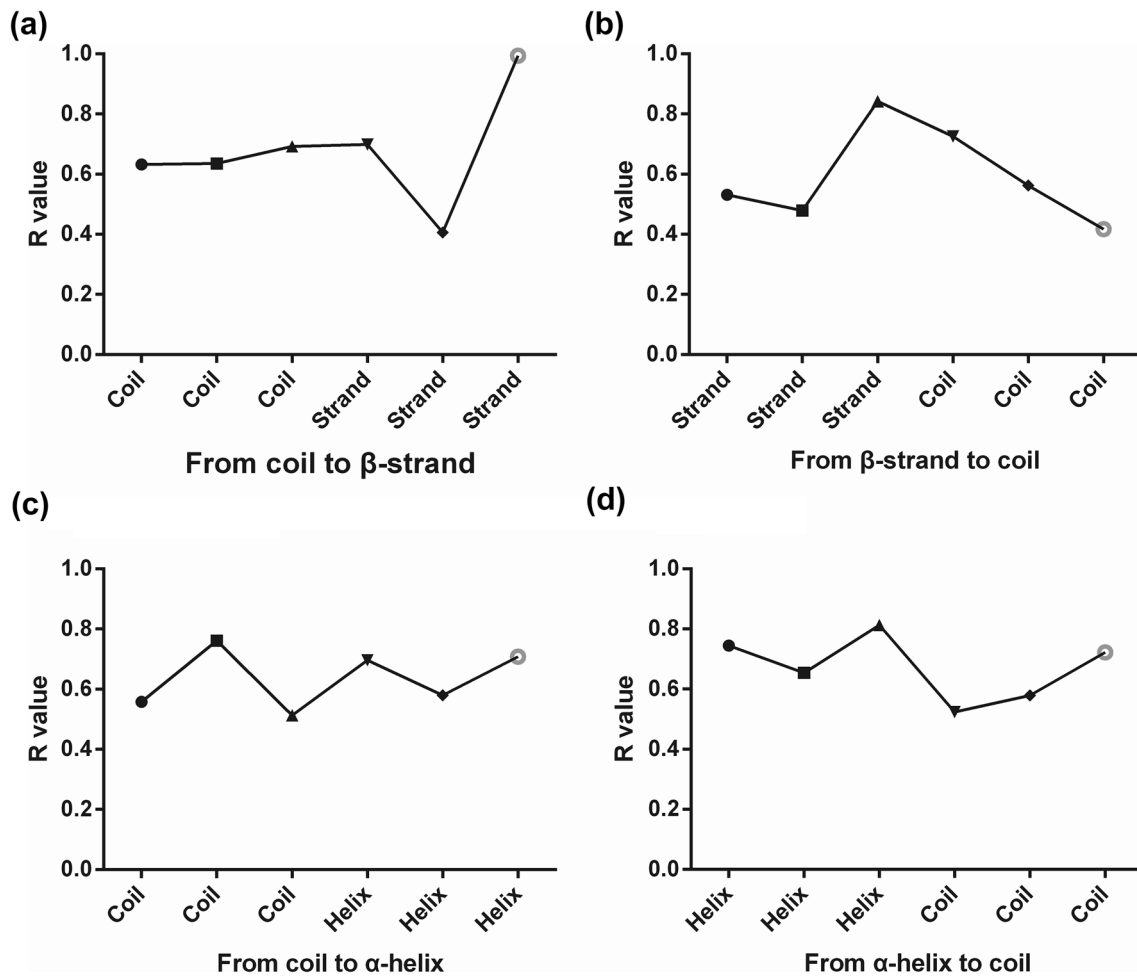


Fig. 2 The changes of R value from one type of folding unit to another in PPRV N protein. **a** Transition boundary from coil to β -strand, x-axis showing the three positions at the C-termination of coil and the three positions at the N-termination of β -strand, **b** transition boundary from β -strand to coil, x-axis showing the three positions at the C-termination of β -strand and the three positions at the

N-termination of coil, **c** transition boundary from coil to α -helix, x-axis showing the three positions at the C-termination of coil and the three positions at the N-termination of α -helix, **d** transition boundary from α -helix to coil, x-axis showing the three positions at the C-termination of α -helix and the three positions at the N-termination of coil

selection of some codons and the specific folding unit (the α -helix, β -strand or coil). Although PPRV was considered as an RNA virus with high mutation rate (Bao et al. 2017), the selection of synonymous codon usages for viral product of PPRV was influenced by translation selection deriving from host (Ma et al. 2017). The overall codon usage bias of N gene reflected by ENC data represented an equivalence relation between the nucleotide changes at the third codon position by mutation pressure and natural selection. To further confirm the influence of natural selection, CAI analysis was frequently used as a measure of gene expression and to their hosts, which reflected the influence of natural selection (Hickey et al. 1995). The higher CAI value is, the more influence of translation selection is in synonymous codon usage bias of the given gene (Carbone et al. 2003; Puigbo et al. 2008b). The prefer fitness of N gene to host at codon

usage reflected translation selection stronger functioned on codon usage of N gene to maintain its normal biological function for adaptation to host. The conserved amino acid sequence of N protein among different PPRV strains enabled the modeling secondary structure to precisely represent the relationship between codon usage patterns and the specific folding unit. As for translation of the virus mRNA sequence, the fine-tuning translation kinetics had obvious effects on the viral product folding (Aragones et al. 2008, 2010). In this study, the synonymous codon usage patterns of the PPRV N gene had a good adaptation to the natural host, and a simple statistical analysis indicated that synonymous codon usage pattern was strongly related to the formation of folding unit. For example, a strong selection of CCU/CCA for Pro existed in the coil unit instead of the α -helix and β -strand units and a biased usage of CGC for Arg existed in the formation

of the β -strand rather than the other two folding units in N protein (Table 1). Our findings were in agreement with other previous reports (Gu et al. 2003; Gupta et al. 2000; Ma et al. 2013; Xie and Ding 1998; Zhou et al. 2013b), suggesting that selection of a specific synonymous codon was maintained on the PPRV N gene for the proper protein structure. Thus, the only target of improvement of exogenous gene in host cell by means of optimal codon replacing rare one in the same synonymous codon family need to be re-taken seriously. Although an agreeable view about the basic features ruling self-folding of proteins into complex secondary structures in in vivo or in vitro is still unclear, translation kinetics might control the co-translational folding pathway and translational pausing at rare codons might cause a time gap to delay the defined portions of the nascent polypeptide emerging from the ribosome (Komar 2009; Kramer et al. 2009). Our study showed that the transition boundaries between the coil unit and the β -strand unit in the PPRV N protein had obvious tendencies to change translation speed (Fig. 2). Translation is physically and functionally combined to the folding and targeting of newly synthesized proteins. Synonymous codon usages and the variations of isoaccepting tRNAs exerted a powerful selective force on translation fidelity and stretches of codons pairing to minor tRNAs form putative sites to locally attenuate translation (Zhang and Ignatova 2009). The adjustment of the translation speed depending on the presence of rare codons in coding sequence may affect the folding efficiency of newly synthesized proteins (Tsai et al. 2008).

In conclusion, the established relationship between the synonymous codon selection and a specific folding unit was based on the biological feature. The synonymous codon pairing to different degrees of tRNA abundance can adjust the translation speed of newly peptide to shape the proper secondary structure of the target protein.

Acknowledgements The work was supported by Central Universities deriving from the Northwest University for Nationalities (31920150077) and Changjiang Scholars and Innovative Research Team in University (IRT_17R88) and National Natural Science foundation of China (No. 31302100).

References

- Aragones L, Bosch A, Pinto RM (2008) Hepatitis A virus mutant spectra under the selective pressure of monoclonal antibodies: codon usage constraints limit capsid variability. *J Virol* 82:1688–1700
- Aragones L, Guix S, Ribes E, Bosch A, Pinto RM (2010) Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog* 6:e1000797
- Bahir I, Fromer M, Prat Y, Linial M (2009) Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* 5:311
- Bailey D, Chard LS, Dash P, Barrett T, Banyard AC (2007) Reverse genetics for peste-des-petits-ruminants virus (PPRV): promoter and protein specificities. *Virus Res* 126:250–255
- Bankamp B, Horikami SM, Thompson PD, Huber M, Billetter M, Moyer SA (1996) Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology* 216:272–277
- Banyard AC, Parida S, Batten C, Oura C, Kwiatek O, Libeau G (2010) Global distribution of peste des petits ruminants virus and prospects for improved diagnosis and control. *J Gen Virol* 91:2885–2897
- Bao J, Wang Q, Li L, Liu C, Zhang Z, Li J, Wang S, Wu X, Wang Z (2017) Evolutionary dynamics of recent peste des petits ruminants virus epidemic in China during 2013–2014. *Virology* 510:156–164
- Baron MD, Diallo A, Lancelot R, Libeau G (2016) Peste des petits ruminants virus. *Adv Virus Res* 95:1–42
- Butt AM, Nasrullah I, Qamar R, Tong Y, 2016, Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg Microbes Infect* 5:e107
- Carbone A, Zinovyev A, Kepes F (2003) Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 19:2005–2015
- Chaney JL, Clark PL (2015) Roles for synonymous codon usage in protein biogenesis. *Annu Rev Biophys* 44:143–166
- Chard LS, Bailey DS, Dash P, Banyard AC, Barrett T (2008) Full genome sequences of two virulent strains of peste-des-petits ruminants virus, the Cote d'Ivoire 1989 and Nigeria 1976 strains. *Virus Res* 136:192–197
- Clarke TF IV, Clark PL (2008) Rare codons cluster. *PLoS ONE* 3:e3412
- Diallo A, Barrett T, Lefevre PC, Taylor WP (1987) Comparison of proteins induced in cells infected with rinderpest and peste des petits ruminants viruses. *J Gen Virol* 68(Pt 7):2033–2038
- Ding YZ, You YN, Sun DJ, Chen HT, Wang YL, Chang HY, Pan L, Fang YZ, Zhang ZW, Zhou P, Lv JL, Liu XS, Shao JJ, Zhao FR, Lin T, Stipkovits L, Pejsak Z, Zhang YG, Zhang J (2014) The effects of the context-dependent codon usage bias on the structure of the nsp1alpha of porcine reproductive and respiratory syndrome virus. *Biomed Res Int* 2014:765320
- dos Reis M, Wernisch L, Savva R (2003) Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res* 31:6976–6985
- Gibbs EP, Taylor WP, Lawman MJ, Bryant J, 1979, Classification of peste des petits ruminants virus as the fourth member of the genus Morbillivirus. *Intervirology* 11:268–274
- Gu W, Zhou T, Ma J, Sun X, Lu Z (2003) Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans Nanobiosci* 2:150–157
- Gu W, Zhou T, Ma J, Sun X, Lu Z (2004) The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems* 73:89–97
- Guisez Y, Robbens J, Remaut E, Fiers W (1993) Folding of the MS2 coat protein in *Escherichia coli* is modulated by translational pauses resulting from mRNA secondary structure and codon usage: a hypothesis. *J Theor Biol* 162:243–252
- Gupta SK, Majumdar S, Bhattacharya TK, Ghosh TC (2000) Studies on the relationships between the synonymous codon usage and protein secondary structural units. *Biochem Biophys Res Commun* 269:692–696
- Guryanov SG, Liljeroos L, Kasaragod P, Kajander T, Butcher SJ (2015) Crystal structure of the measles virus nucleoprotein core in complex with an N-terminal region of phosphoprotein. *J Virol* 90:2849–2857
- Gutsche I, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RW, Sachse C, Schoehn G (2015) Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* 348:704–707

- Hanson G, Collier J (2017) Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol* 19:20–30
- Hickey PL, Angus PW, McLean AJ, Morgan DJ (1995) Oxygen supplementation restores theophylline clearance to normal in cirrhotic rats. *Gastroenterology* 108:1504–1509
- Ismail TM, Yamanaka MK, Saliki JT, el-Kholy A, Mebus C, Yilma T (1995) Cloning and expression of the nucleoprotein of peste des petits ruminants virus in baculovirus for use in serological diagnosis. *Virology* 208:776–778
- Kanduc D (2017) Rare human codons and HCMV translational regulation. *J Mol Microbiol Biotechnol* 27:213–216
- Komar AA (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* 34:16–24
- Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* 16:589–597
- Liu F, Wu X, Zhao Y, Li L, Wang Z (2014) Budding of peste des petits ruminants virus-like particles from insect cell membrane based on intracellular co-expression of peste des petits ruminants virus M, H and N proteins by recombinant baculoviruses. *J Virol Methods* 207:78–85
- Ma XX, Feng YP, Liu JL, Ma B, Chen L, Zhao YQ, Guo PH, Guo JZ, Ma ZR, Zhang J (2013) The effects of the codon usage and translation speed on protein folding of 3D(pol) of foot-and-mouth disease virus. *Vet Res Commun* 37:243–250
- Ma XX, Chang QY, Ma P, Li LJ, Zhou XK, Zhang DR, Li MS, Cao X, Ma ZR (2017) Analyses of nucleotide, codon and amino acids usages between peste des petits ruminants virus and rinderpest virus. *Gene* 637:115–123
- Marin M (2008) Folding at the rhythm of the rare codon beat. *Biotechnol J* 3:1047–1057
- Mioduser O, Goz E, Tuller T (2017) Significant differences in terms of codon usage bias between bacteriophage early and late genes: a comparative genomics analysis. *BMC Genom* 18:866
- Mitra-Kaushik S, Nayak R, Shaila MS (2001) Identification of a cytotoxic T-cell epitope on the recombinant nucleocapsid proteins of Rinderpest and Peste des petits ruminants viruses presented as assembled nucleocapsids. *Virology* 279:210–220
- Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 28:292
- Nossmann M, Pieper J, Hillmann F, Brakhage AA, Munder T (2017) Generation of an arginine-tRNA-adapted *Saccharomyces cerevisiae* strain for effective heterologous protein expression. *Curr Genet*. <https://doi.org/10.1007/s00294-017-0774-8>
- Parida S, Muniraju M, Mahapatra M, Muthuchelvan D, Buczkowski H, Banyard AC (2015) Peste des petits ruminants. *Vet Microbiol* 181:90–106
- Pouyet F, Mouchiroud D, Duret L, Semon M (2017) Recombination, meiotic expression and human codon usage. *Elife* 6:e27344
- Puigbo P, Bravo IG, Garcia-Vallve S (2008a) CALcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* 3:38
- Puigbo P, Bravo IG, Garcia-Vallve S (2008b) E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* 9:65
- Rafels-Ybern A, Torres AG, Grau-Bove X, Ruiz-Trillo I, Ribas de Pouplana L (2017) Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two bacterial phyla. *RNA Biol* 2017:1–8
- Rodriguez A, Wright G, Emrich S, Clark PL (2017) %MinMax: a versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding. *Protein Sci* 27:356–362
- Sharp PM, Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15:1281–1295
- Singh VK, Kumar V, Krishnamachari A (2017) Prediction of replication sites in *Saccharomyces cerevisiae* genome using DNA segment properties: multi-view ensemble learning (MEL) approach. *Biosystems* 163:56–69
- Straube R (2017) Analysis of network motifs in cellular regulation: structural similarities, input-output relations and signal integration. *Biosystems* 162:215–232
- Tsai CJ, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R (2008) Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol* 383:281–291
- Welch M, Villalobos A, Gustafsson C, Minshull J (2009) You're one in a googol: optimizing genes for protein expression. *J R Soc Interface* 6(Suppl 4):S467–S476
- Weygand-Durasevic I, Ibba M (2010) Cell biology. New roles for codon usage. *Science* 329:1473–1474
- Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87:23–29
- Xie T, Ding D (1998) The relationship between synonymous codon usage and protein structure. *FEBS Lett* 434:93–96
- Zhang G, Ignatova Z (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS ONE* 4:e5036
- Zhou JH, Gao ZL, Zhang J, Ding YZ, Stipkovits L, Szathmari S, Pejsak Z, Liu YS (2013a) The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infect Genet Evol* 14:105–110
- Zhou JH, You YN, Chen HT, Zhang J, Ma LN, Ding YZ, Pejsak Z, Liu YS (2013b) The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus. *Infect Genet Evol* 16:270–274