Infection, Genetics and Evolution 16 (2013) 270–274

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

The effects of the synonymous codon usage and tRNA abundance on protein folding of the 3C protease of foot-and-mouth disease virus

Jian-hua Zhou [a,1], Ya-nan You [a,1], Hao-tai Chen [a,1], Jie Zhang [a], Li-na Ma [a], Yao-zhong Ding [a], Zygmunt Pejsak [b], Yong-sheng Liu [a,*]

[a] 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, PR China
[b] Department of Swine Diseases, National Veterinary Research Institute, 57 Partyzantow, 24-100 Puławy, Poland

ARTICLE INFO

Article history:
Received 28 November 2012
Received in revised form 18 February 2013
Accepted 22 February 2013
Available online 14 March 2013

Keywords:
Foot-and-mouth disease virus
3C protease
Synonymous codon usage
Folding unit
Translation speed
tRNA abundance

ABSTRACT

The 3C protease of foot-and-mouth disease virus (FMDV) has a conserved amino acid sequence and is responsible for most cleavage in the viral polyprotein. The effects of the synonymous codon usage of FMDV 3C gene and tRNA abundance of the hosts on shaping different folding units (α-helix, β-strand and the coil) in the 3C protease were analyzed based on the structural information of the FMDV 3C protease from Protein Data Bank (PDB: 2BHG) and 210 genes of 3C for all serotypes of FMDV. The strong correlation between some codons usage and the specific folding unit in the FMDV 3C protease is found. As for the effect of translation speed caused by tRNA abundance on the formation of folding units, the codon positions with lowly abundant tRNA scatter in the 3C gene and there is the obvious fluctuation of tRNA abundance locating in the transition boundaries from the β-strand to the α-helix and the coil, respectively. However, codon positions with lowly abundant tRNA clustering into these boundaries are not found, suggesting that the adjustment of translation speed is likely also achieved by the single codon position with low tRNA abundance rather than a cluster. The observations can provide the information for insight into the role of the synonymous codon usage in the formation of 3C protease of FMDV and effect of the tRNA abundance of the hosts on this formation of protease.

1. Introduction

Foot-and-mouth disease virus (FMDV) is a member of the *aphthovirus* genus of the *picornavirus* family, which is one of important group of mammalian single-strand and positive-sense RNA virus (Mason et al., 2003). As for the infection caused by FMDV, the range of susceptible animal is relatively wide, such as pig, cattle and sheep etc. cloven-hoofed animals, and guinea pig is a major animal model for FMDV infection research. There are many factors for the pathogenicity, one of the factors is that the replication and translation strategy requires the translation of a polyprotein precursor. The precursor that can be efficiently cleaved by virally encoded protease into various mature proteins which shape the viral capsid and cater to replication mechanism (Mason et al., 2003). The conserved 3C protease is a very important segment in this process and is expected to serve as an available anti-viral drug target (Curry et al., 2007a; Curry et al., 2007b). The crystal structure of FMDV 3C protease shows that, in common with other picornaviral 3C proteases, it adopts a conserved secondary structure similar to that of the archetypal serine protease chaymotrypsin (Birtley et al., 2005). When it comes to the catalytic mechanism of FMDV 3C protease, the previous studies with *Escherichia coli*-expressed mutant forms of 3C protease showed that Cys_{161}, His_{46} and Asp_{84} make up of the catalytic triad (Grubman et al., 1995). There has been many studies on the genetic diversity, the structure and catalytic sites of FMDV 3C protease (Belsham et al., 2000; Birtley et al., 2005; Biswas et al., 2006; Carrillo et al., 2005), but little information about a relationship between synonymous codon usage and the secondary structure of FMDV 3C protease. Up to date, the synonymous codon usage pattern has been found to be linked with many factors, such as genetic diversity, translation efficiency, transfer RNA, amino acid conservation and protein structure, etc. (Bahir et al., 2009; Clarke and Clark, 2008; dos Reis et al., 2003; Welch et al., 2009; Weygand-Durasevic and Ibba, 2010). The synonymous codon mutation can so profoundly change the function of a protein and add a new level of complexity to how the genetic code is interpreted (Weygand-Durasevic and Ibba, 2010). There is a high correlation between translation efficiency and the synonymous codon usage pattern (Ikemura, 1985; Li and Luo, 1996; Sorensen et al., 1989). As for the formation of protein secondary structures, mRNA sequences have in principle an additional potential to carry

* Corresponding author. Tel.: +86 931 8342771; fax: +86 931 8340977.
E-mail address: liuyongshengvip8@163.com (Y.-s. Liu).
[1] These authors have equal contribution to the study.

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.02.017

structural information concerning the encoded protein due to the degeneracy of the genetic code, which can be at the level of a single codon or a nucleotide fragment (Guisez et al., 1993; Xie and Ding, 1998). It is noted that one synonymous codon can correspond to the specific tRNA abundance(Miyasaka, 2002). This feature can be explained by the fact that, when the pools of their corresponding tRNAs are large, the waiting time till correct tRNA attachment at the ribosome is reduced, thus lowering the chance for incorrect tRNA attachment (Bennetzen and Hall, 1982; Ikemura, 1982). The experimental results and theoretical *in silico* modeling studies provide evidence that mRNA might contain an additional layer of information, beyond the amino acid sequence, that fine-tunes *in vivo* protein folding, and the fine-tunes are largely believed to occur in a co-translational process (Komar, 2009). As for translation of the virus mRNA sequence, the fine-tuning translation kinetics might also mediate the viral product folding (Aragones et al., 2008; Aragones et al., 2010). Although the relationship between the synonymous codon usage patterns and formation of many protein secondary structures has been studied to a large degree (Saunders and Deane, 2010; Weygand-Durasevic and Ibba, 2010; Xie and Ding, 1998; Zhou et al., 2009), the relationship is not investigated in viruses. The secondary structure of FMDV 3C protease is available from Protein Data Bank and 3C is conserved during evolution of FMDV. The information facilitates the analysis for potential relationship between synonymous codon usage and the secondary structure of FMDV 3C. Furthermore, the effects of tRNAs abundance of the three hosts on the formation of secondary structure of 3C were analyzed in this study.

2. Materials and methods

2.1. Information of sequence and structure of FMDV 3C protease

The 210 coding sequences, including 153 sequences coding for the FMDV polyprotein and 57 sequences coding for the FMDV 3C protease was downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/) and the detailed information is listed in Table S1. To investigate the synonymous codon usage pattern of the 3C protease, the coding sequence for the FMDV 3C protease were obtained from the whole coding sequence by multiple sequence alignments performed with the Clustal W (1.7) computer programs (Thompson et al., 1994). The information about the secondary structure of the FMDV 3C protease was obtained from Protein Data Bank (PDB: 2BHG). Additionally, to analyze the tRNA abundance corresponding to each codon position along the coding sequence for the 3C protease, the copy numbers of the cognate tRNAs in cattle, sheep and guinea pig were obtained from tRNA database (http://lowelab.ucsc.edu/) (Table S2).

2.2. Analysis of the overall tRNA abundance for each codon position of 3C gene

To identify the translation selection, caused by the various tRNA copy numbers, at each codon position in the FMDV 3C gene, we devised an index (R value) representing the overall tRNA abundance for a particular codon position in a target gene.

$$R = \sqrt[n]{\prod_{1}^{n}(Wij/Wj)}$$

where R value indicates that the overall tRNA abundance for a particular codon position in the interesting 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, n means the number of the interesting gene. The R value ranges from 0.00 to 1.00. The R value less than 0.3 for a codon position is defined to be recognized by lowly abundant tRNAs, and the R value more than 0.7 for a codon position indicates this position is recognized by high abundant tRNAs.

2.3. Analysis of relationship between the synonymous codon usage and the structure of 3C protease

Based on the alignment between the amino acid sequences of the FMDV 3C protease (PDB: 2BHG) and the 210 sequences of the FMDV 3C protease involved in this study, we can locate the different secondary structure units in the target protein and focus on tRNA abundance in the boundaries comprising of the C-termination of the upstream folding unit and the N-termination of the downstream unit. To estimate specific correlations between the synonymous codon usage and the secondary units of the interesting protein, we devised a formula to calculate P value.

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

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

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

where $N_{(i,sec\text{-}k)}$ 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 is much greater than 1.5, the synonymous codon has a strong tendency to exist in the interesting secondary unit; when P value is much less than 0.5, the synonymous codon has a strong tendency to free from the interesting secondary unit.

3. Result

3.1. The overall tRNA abundance for each codon position of 3C gene

Base on R data, we mapped to describe the tRNA abundance for each codon position along the FMDV 3C gene. We assume that the translation speed for the synthesis of FMDV 3C protease is not stable in the hosts, since the fluctuation of tRNA abundance of the three hosts for each codon position along the 3C gene was found (Fig. S1). The codon positions with R value much less than 0.30 have little chance to cluster in the FMDV 3C gene in the three animal cells, expect for positions 1–3. However, the codon positions including position 7–8, 21–24, 30–31, 41–44, 47–48, 55–58, 66–69, 79–82, 87–91, 120–121, 138–139, 148–149, 176–177, 190–191, 198–199, 205–209, 211–212 with R value much greater than 0.7 tend to cluster in the FMDV 3C gene among the three animal cells. The kind of genetic feature do not result in ending the expression of the FMDV 3C gene, but assists ribosomes into scanning this gene and synthesizing the target product successfully.