



Mapping codon usage in sequence regions flanking cleavage positions in the hepatitis A virus polyprotein

X.-X. Ma*, Y.-P. Feng*, L. Chen, Y.-Q. Zhao, J.-L. Liu, J.-Z. Guo, P.-H. Guo, J.-T. Yang, J.-X. Lu, S.-E. Chen and Z.-R. Ma

College of Life Science and Engineering,
Northwest University for Nationalities, Lanzhou, China

*These authors contributed equally to this study.

Corresponding author: Z.-R. Ma

E-mail: mazhongren@xbmu.edu.cn / maxiaoxia956@163.com

Genet. Mol. Res. 12 (3): 2306-2319 (2013)

Received December 10, 2012

Accepted May 14, 2013

Published July 8, 2013

DOI <http://dx.doi.org/10.4238/2013.July.8.11>

ABSTRACT. To analyze the synonymous codon usage patterns of sequence regions flanking cleavage sites in the hepatitis A virus (HAV) polyprotein, the codon usage bias at codon positions and the synonymous codon usage in the target contexts of 30 virus strains were estimated by two simple methods that were based on the values for relative synonymous codon usage. In addition, the pattern of synonymous codon usage was compared between the genomic sequences in HAV and those of its human host. Our results indicated that HAV adopts a combination of coincidence and antagonism with the synonymous codon usage in humans. This characteristic may help HAV to efficiently use the translational machinery in its human host. We also observed that codon usage exhibited a strong bias in some specific positions in these contexts, and that the underrepresented synonymous codons, CUA for Leu, ACG for Thr, GUA for Val, and UCG for Ser, are preferentially used in these positions. These underrepresented synonymous codons likely play roles in regulating the rate of protein translation and influencing the secondary

structure of the sequence regions flanking the cleavage sites.

Key words: Hepatitis A virus; Cleavage site, Codon usage bias; Human host; Synonymous codon usage

INTRODUCTION

The hepatitis A virus (HAV) is a member of the Hepatovirus genus belonging to the Picornaviridae family (Pintó et al., 2007). HAV is a non-enveloped icosahedral virus with a positive single-stranded RNA genome. The genome contains an open reading frame (ORF) encoding a polyprotein, which is flanked by 5' and 3' non-coding regions (Cohen et al., 1987). After translation, this polyprotein is subsequently cut into 11 proteins through a cascade of proteolytic events performed mainly by the viral 3C protease (Schultheiss et al., 1994, 1995). Although many reports have indicated that the overall genomic organization and the translation strategy of HAV are highly similar to those in other members of Picornaviridae family, many differences still exist. These include differences in the overall codon usage pattern of HAV (Pintó et al., 2007; Sablok et al., 2011) and the effects of synonymous codon usage pattern of host and virus on expression of the HAV genome (Sánchez et al., 2003; Pintó et al., 2007). Previous studies have shown that the genomes of human RNA viruses have high mutation rates, caused mainly by mutation pressure (Drake and Holland, 1999). The above-mentioned studies focused mainly on the general codon usage pattern of RNA viruses to identify the nature of the mutation pressures or the mechanisms of translation selection during the evolution of RNA viruses. Comparative analysis of the synonymous codon usage plays an essential role in investigating viral evolution, particularly for comparisons of the synonymous codon usage between RNA viruses and their hosts (Barrai et al., 2008; Bahir et al., 2009; Wong et al., 2010; Zhou et al., 2012). In general, three main codon usage patterns between RNA viruses and their natural hosts have been identified: antagonism, a mixture of coincidence and antagonism, and coincidence. It is clear that comparative analysis of the overall codon usage pattern between RNA viruses and their hosts can reveal coevolution and adaptation of viruses to their hosts. Such information is also valuable for investigating the strategy of viral gene expression. Efforts to identify the translational feature of capsid-coding sequence in HAV, involving comparative analysis of the synonymous codon usage pattern in the virus and its human host, have indicated that antagonistic codon usage appears to be required for the correct formation of the viral capsid structure (Sánchez et al., 2003; Pintó et al., 2007). Moreover, several studies have reported that the usage bias in synonymous codons in some ORF regions may regulate the efficiency of expression from these regions (Kane, 1995; Zolotukhin et al., 1996; Ohno et al., 2001; Song et al., 2004; Zhi et al., 2010; Zhou et al., 2011, 2013a,b). These biological phenomena may be the result of selection for fine-tuned translation kinetics for regulating expression efficiency (Aragonès et al., 2008, 2010). For the formation of packaging virions, the precise cleavage of the HAV precursor polyprotein plays an important role because this cleavage is a key step in virion formation (Buenz and Howe, 2006). However, little information about the codon usage in the sequence regions flanking the cleavage sites in the HAV polyprotein is available to date. Here, we used 30 strains of HAV as a model to analyze the characteristics of codon usage in the contexts flanking the cleavage sites in the viral polyprotein.

MATERIAL AND METHODS

Sequence information

To avoid repeated retrieval of sequences with high homology, 30 complete RNA sequences of HAV were selected and downloaded from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>) (Table S1). To locate the cleavage sites in the complete coding sequence of HAV, multiple sequence alignments were performed with the ClustalW (1.7) method of the DNASTAR software (7.0) for Windows, with minor manual editing.

Calculation of relative synonymous codon usage (RSCU)

The RSCU values for the 30 coding sequences of HAV were calculated as previously described (Sharp and Li, 1986). RSCU values do not depend on amino acid composition and the size of the coding sequence, because these 2 variables can be eliminated in the calculation. An RSCU value for a codon equal to 1.0 indicates that this codon is chosen equally and randomly. The RSCU value for a synonymous codon >1.0 or <1.0 indicates a greater or lower frequency of this codon, respectively. Synonymous codons with RSCUs >1.6 were considered to be overrepresented, whereas synonymous codons with RSCUs <0.6 were regarded as underrepresented (Wong et al., 2010). In addition, in order to analyze the synonymous codon usage pattern between HAV and its natural host, the frequency of synonymous codon usage in the human genome was obtained from the codon usage database <http://www.kazusa.or.jp/codon/> (Nakamura et al., 2000) and calculated by the RSCU formula. The RSCU data were then used to identify any differences in synonymous codon usage between HAV and human.

Calculation of the codon bias model for each position flanking the cleavage sites of the HAV polyprotein

To investigate the codon usage model for each position flanking the cleavage sites of the HAV polyprotein, a formula for CUB_i was devised to evaluate the accumulation of codon usage bias at each position flanking the target contexts of the HAV polyprotein.

$$CUB_i = \sqrt[30]{\prod_{k=1}^{k=30} RSCU_{ki}}$$

where the CUB_i value represents the codon bias at a certain position in the target contexts in HAV; the k value represents 30 sequences of HAV; the i value represents a specific position ranging from 1-9 and 12-20 up- and downstream of each cleavage site (positions 10 and 11), respectively. A CUB_i value <0.6 indicated that the codon usage had a low bias at this position in the examined region; conversely, a CUB_i value >1.6 indicated highly biased codon usage at this position.

Calculation of the synonymous codon usage in the contexts of flanking cleavage sites of the HAV polyprotein

To calculate the discrepancy in synonymous codon usage between the target contexts

and the whole coding sequence of HAV, we developed a simple method based on a previous report (Zhou et al., 2011).

$$R = \ln\left(\frac{n_{20} / N_{20}}{n / N}\right)$$

where n_{20} is the sum of synonymous codons in the 20 codons flanking the cleavage site, N_{20} is the sum of the corresponding amino acids in this 20-codon region, n is the sum of synonymous codons in the entire coding sequence, and N is the sum of the corresponding amino acids in the entire coding sequence.

RESULTS

Comparative analysis of the RSCU values between HAV and human sequences

For the synonymous codon usage pattern of the HAV sequence, the most overrepresented synonymous codons were those ending in A or U (except for UUG for Leu), including UUU for Phe, AUU for Ile, GUU for Val, UCU and UCA for Ser, CCU and CCA for Pro, ACU and ACA for Thr, GCU for Ala, CAU for His, AAU for Asn, GAU for Asp, UGU for Cys, and AGA for Arg. On the other hand, the most underrepresented synonymous codons were those ending in C or G (except for CUA for Leu, GUA for Val, and CGA for Arg), including UUC for Phe; CUC for Leu; AUC for Ile; GUC for Val; UCG and AGC for Ser; CCC and CCG for Pro; ACC and ACG for Thr; GCG for Ala; UAC for Tyr; CAC for His; AAC for Asn; GAC for Asp; UGC for Cys; CGU, CGC, and CGG for Arg; and GGC for Gly (Table 1). This observed codon usage bias in the HAV genome suggests that some synonymous codons are not chosen equally and randomly.

We noted a mixture of coincidence and antagonism of synonymous codon usage pattern between HAV and its natural host. In particular, similar synonymous codon usage pattern included CUA for Leu, AUA for Ile, GUA for Val, UCG and AGU for Ser, CCG for Pro, ACG for Thr, GCG and GCA for Ala, CAA and CAG for Glu, AAA and AAG for Glu, and GGG for Gly. In contrast, the reversed synonymous codon usage pattern included UUA, UUG, CUC, and CUG for Leu; AUU and AUC for Ile; GUU and GUC for Val; UCU, UCA, and AGC for Ser; CCC for Pro; ACU and ACC for Thr; GCU and GCC for Ala; UAC for Tyr; CAC for His; AAU and AAC for Asn; GAU and GAC for Asp; UGU and UGC for Cys; CGC, CGG, and AGA for Arg; and GGC and GGA for Gly (Table 1). Among the synonymous codons with similar usage, a similar underrepresentation of synonymous codons is found for CUA, GUA, UCG, CCG, ACG, and GCG codons, which - in both HAV and human sequences - contain TpA or CpG at their 3' ends (Table 1). These observations likely suggest that these codons regulate the speed of translation from sequence regions in which they reside. The underrepresentation of CpG or TpA dinucleotides has also been found in several other organisms (Cooper and Youssoufian, 1988; Karlin and Burge, 1995; De Amicis and Marchetti, 2000; Sugiyama et al., 2005).

The codon usage bias at each position flanking every cleavage site in HAV

In general, the degree of codon usage bias was highly variable in the regions flanking the

Table 1. Relative synonymous codon usage (RSCU) data of hepatitis A virus (HAV) and human.

Amino acid	Codon	RSCU ^a	RSCU ^b
Phe	UUU	1.570	0.871
	UUC	0.430	1.129
Leu	UUA	1.300	0.388
	UUG	2.490	0.731
	CUU	1.140	0.729
	CUC	0.190	1.215
	CUA	0.230	0.404
Ile	CUG	0.650	2.533
	AUU	2.040	1.035
	AUC	0.320	1.522
	AUA	0.640	0.443
Val	GUU	2.240	0.691
	GUC	0.320	0.996
	GUA	0.390	0.416
	GUG	1.050	1.897
Ser	UCU	2.340	1.107
	UCC	0.660	1.394
	UCA	2.020	0.836
	UCG	0.120	0.330
	AGU	0.740	0.836
	AGC	0.120	1.498
Pro	CCU	1.920	1.121
	CCC	0.410	1.354
	CCA	1.620	1.066
	CCG	0.050	0.460
Thr	ACU	1.820	0.938
	ACC	0.360	1.524
	ACA	1.720	1.074
	ACG	0.100	0.463
Ala	GCU	2.120	1.091
	GCC	0.610	1.640
	GCA	1.250	0.852
	GCG	0.030	0.418
Tyr	UAU	1.580	0.839
	UAC	0.420	1.161
His	CAU	1.660	0.809
	CAC	0.340	1.191
Gln	CAA	0.950	0.507
	CAG	1.050	1.493
Asn	AAU	1.630	0.890
	AAC	0.370	1.110
Lys	AAA	1.270	0.822
	AAG	0.730	1.178
Asp	GAU	1.660	0.891
	GAC	0.340	1.109
Glu	GAA	1.120	0.813
	GAG	0.880	1.187
Cys	UGU	1.650	0.858
	UGC	0.350	1.142
Arg	CGU	0.160	0.512
	CGC	0.120	1.196
	CGA	0.130	0.633
	CGG	0.030	1.196
	AGA	4.320	1.202
	AGG	1.240	1.260
Gly	GGU	1.100	0.642
	GGC	0.500	1.397
	GGA	1.780	0.977
	GGG	0.620	0.984

^aRSCU values were calculated based on the synonymous codon usage frequencies of HAV. ^bRSCU values were calculated based on the synonymous codon usage frequencies of human.

cleavage sites of the HAV polyprotein, and these biases (indicated by *CUB* values <0.6 or >1.6) affected specific positions in these regions. In particular, a high bias (*CUB* value >1.6) of codon usage was detected at positions 8, 9, 12, and 18 in the VP4/VP2 context; positions 2, 5, 8, 9, 13, 14, and 17 in VP2/VP3; positions 3, 4, 6, 7, 9, 11, 12, 14, 15, and 19 in VP3/VP1; positions 1-3, and 8 in VP1/2A; positions 1, 2, 7, 12-14, 19, 20 in 2A/2B; positions 2, 8, 9, 17, and 20 in 2B/2C; positions 7, 9, 11, 13, and 14 in 2C/3A; positions 9, 14, and 17 in 3A/3B; positions 6, 9, 11, 12, 19, and 20 in 3B/3C; and positions 3, 7, 11, 18, and 20 in the 3C/3D context (Figures S1-S10). The amino acids Ile, His, Thr, Ser, Arg, Asn, Ala, Asp, Val, Gly, and Pro were preferentially used in these positions. Moreover, a low bias of codon usage was found at positions 4-6 in the VP4/VP2 context, positions 1 and 11 in 2B/2C, position 1 in 3A/3B, and position 7 in the 3B/3C context (Figures S1, S6, S8, S9). The amino acids Asp, His, Ile, Ser, and Val were preferentially used in these positions, and their favored synonymous codons tended to be chosen at these positions.

Bias of synonymous codon usage in the given contexts flanking the cleavage sites

Next, we investigated the association between synonymous codon usage and the codon usage bias in positions flanking each cleavage site in the HAV polyprotein. Some synonymous codons with RSCU values <0.6 were preferentially used in the regions flanking the cleavage sites in the HAV protein (Figures 1-10). The synonymous codons with *R* values >0 and RSCU values <0.6 (calculated from the codon usage in human DNA sequences) in the various regions flanking the cleavage sites of HAV are listed in Table 2. These codons were of two types: one type was represented by UUA for Leu, CAA for Gln, and AUA for Ile whose RSCU values, calculated from the whole coding sequence of HAV, were not coincident with those in human sequences. The other codon type was represented by CUA for Leu, GUA for Val, UCG for Ser, CCG for Pro, and CGU for Arg, all of which were underrepresented codons and encoded hydrophobic amino acids in both HAV and human sequences (Tables 1 and 2).

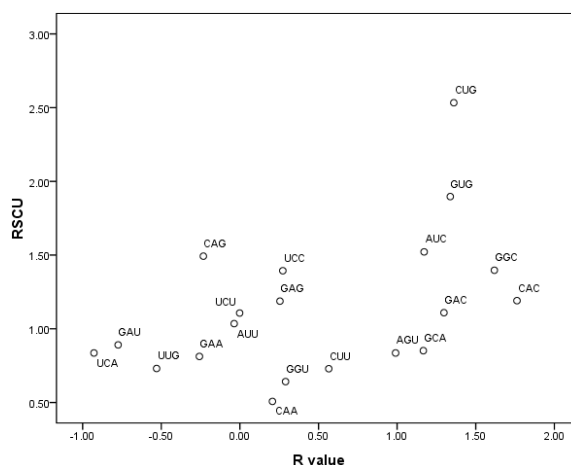


Figure 1. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the VP4/VP2 of the hepatitis A virus (HAV). Each plot represents the usage degree of a synonymous codon, which exists in the target context in the present study. When *R* value >0 , the synonymous codon is more preferentially used in the target context than in the whole coding sequence of HAV. RSCU = relative synonymous codon usage.

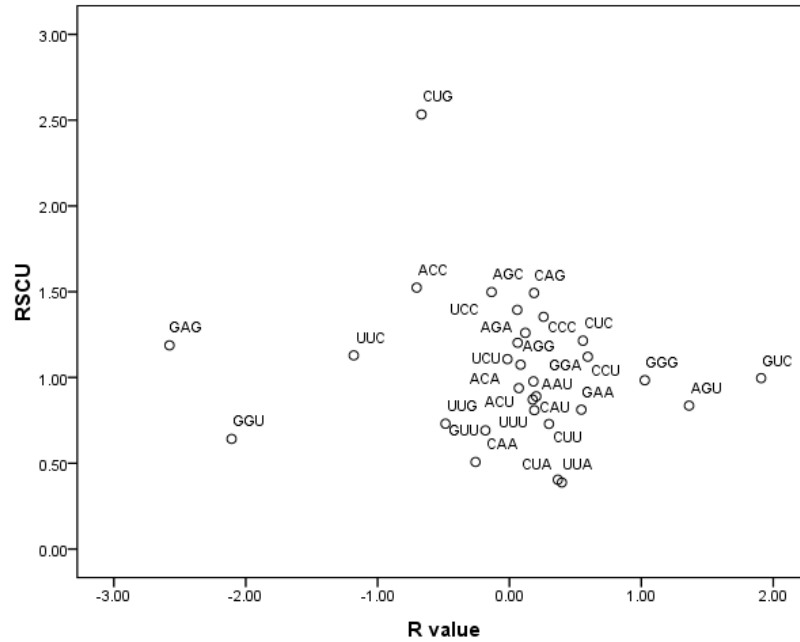


Figure 2. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the VP2/VP3 of the hepatitis A virus. RSCU = relative synonymous codon usage.

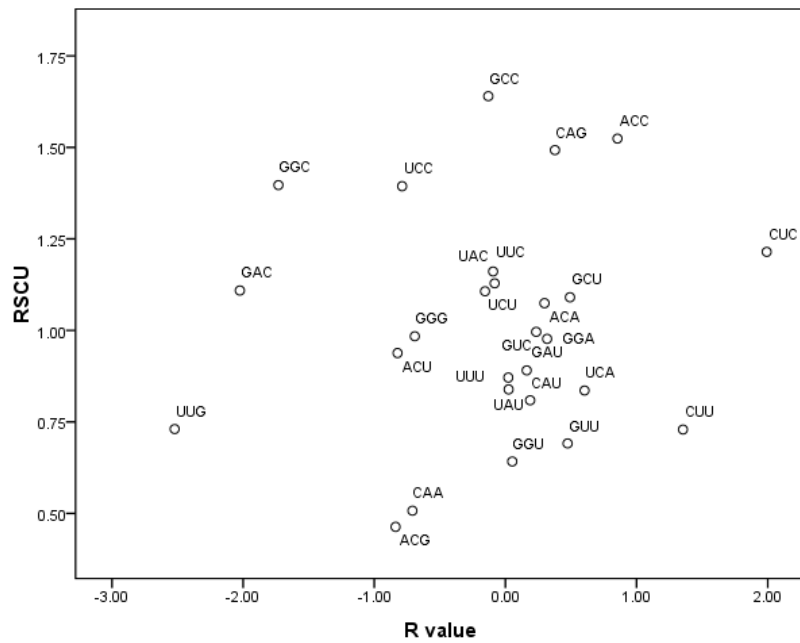


Figure 3. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the VP3/VP1 of the hepatitis A virus. RSCU = relative synonymous codon usage.

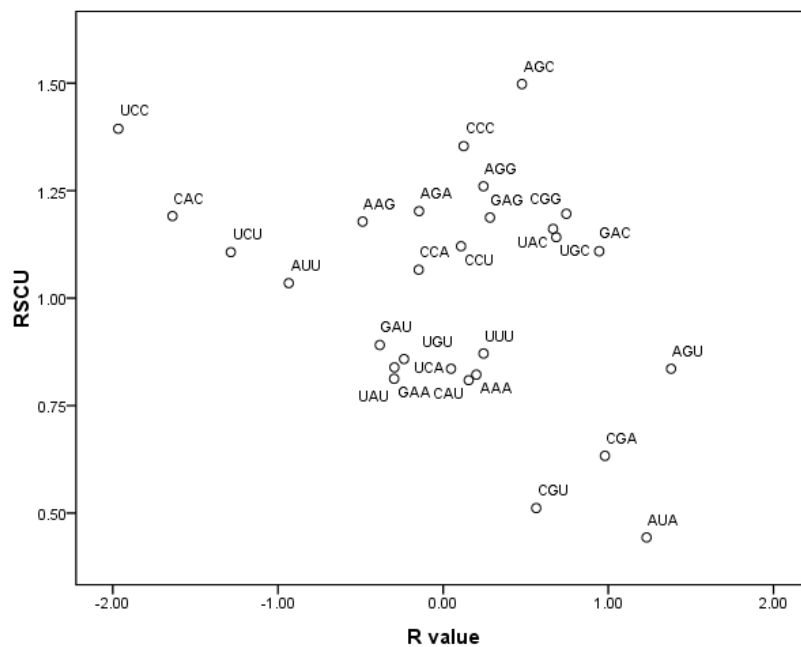


Figure 4. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the VP1/2A of the hepatitis A virus. RSCU = relative synonymous codon usage.

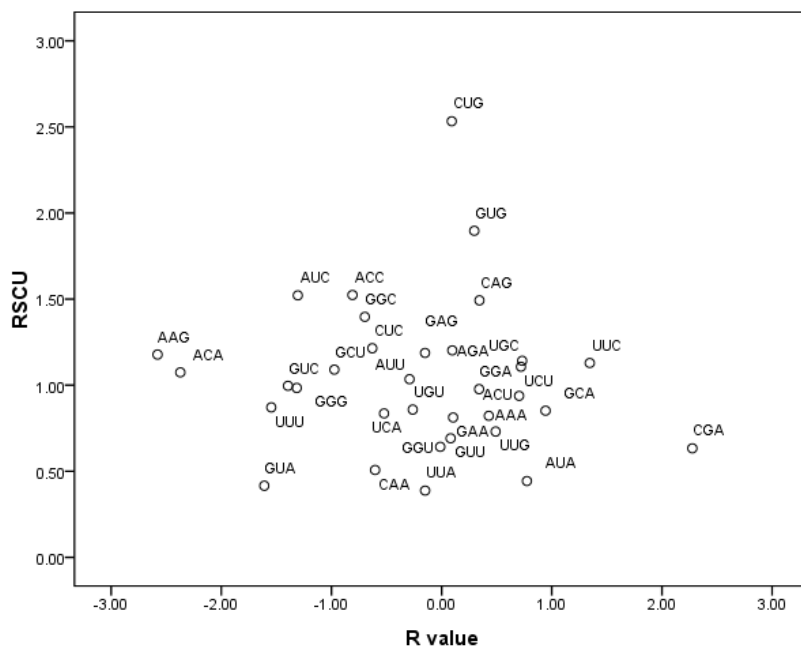


Figure 5. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 2A/2B of the hepatitis A virus. RSCU = relative synonymous codon usage.

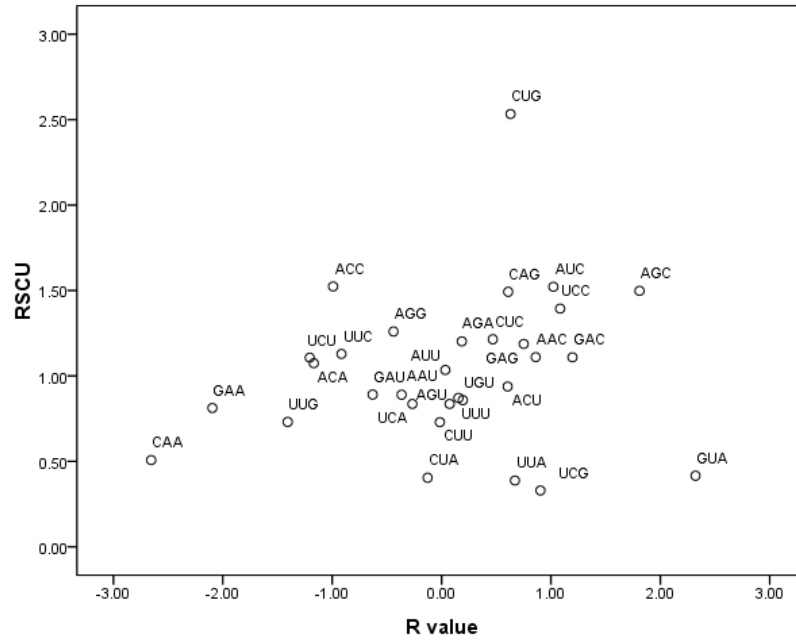


Figure 6. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 2B/2C of the hepatitis A virus. RSCU = relative synonymous codon usage.

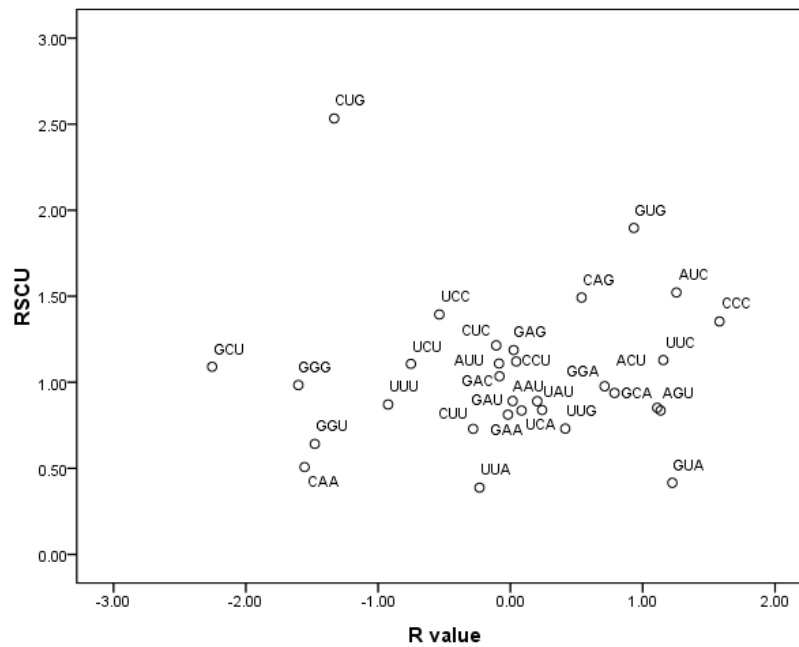


Figure 7. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 2C/3A of the hepatitis A virus. RSCU = relative synonymous codon usage.

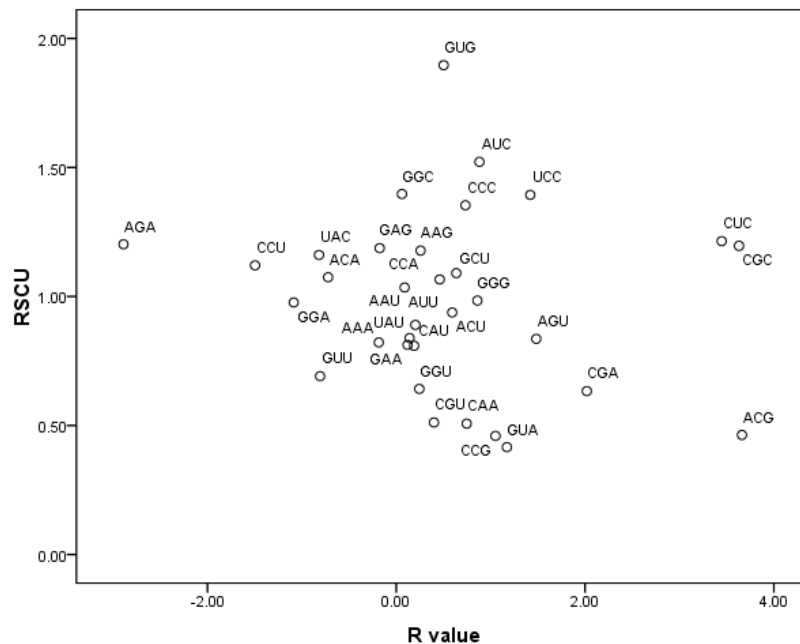


Figure 8. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 3A/3B of the hepatitis A virus. RCSU = relative synonymous codon usage.

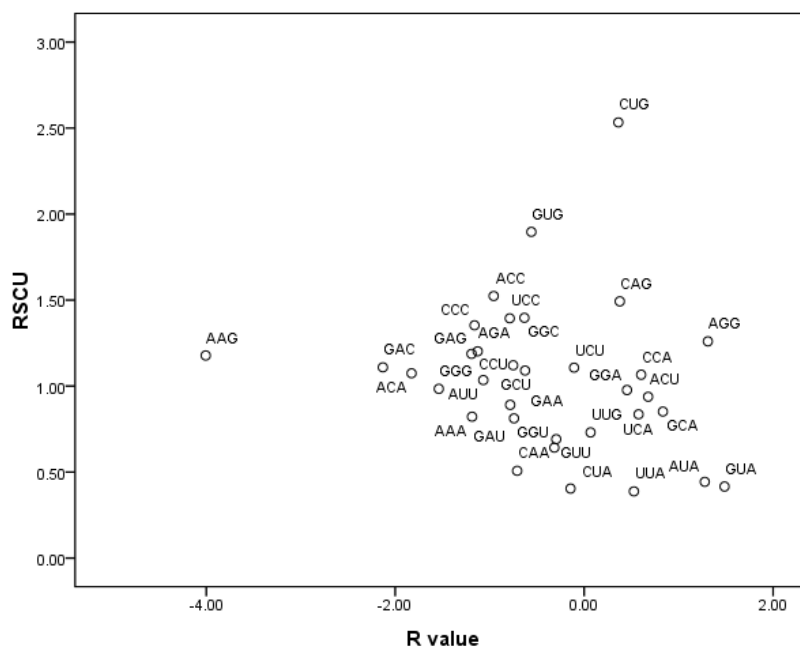


Figure 9. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 3B/3C of the hepatitis A virus. RCSU = relative synonymous codon usage.

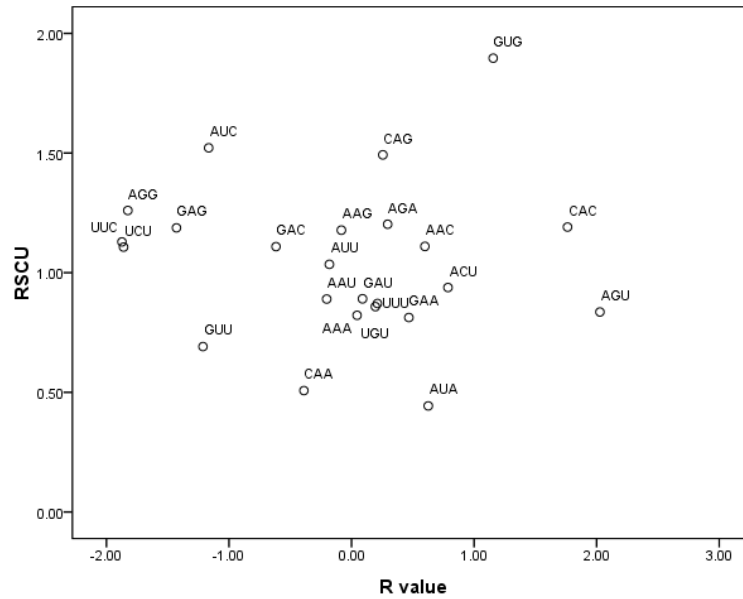


Figure 10. Relationship between the synonymous codon usage and the codon usage bias of the positions flanking the 3C/3D of the hepatitis A virus. RSCU = relative synonymous codon usage.

Table 2. Underrepresented codons for human preferentially selected in positions flanking the cleavage sites of the hepatitis A virus.

Cleavage site	Codon/amino acid	R (>0)	*RSCU (<0.6)
VP4/VP2	CAA / Gln	0.207	0.507
VP2/VP3	UUA / Leu	0.397	0.388
	CUA / Leu	0.367	0.404
VP1/2A	AUA / Ile	1.232	0.443
	CGU / Arg	0.564	0.512
2A/2B	AUA / Ile	0.774	0.443
2B/2C	UUA / Leu	0.671	0.388
	GUA / Val	2.322	0.416
	UCG / Ser	0.904	0.330
2C/3A	GUA / Val	1.223	0.416
3A/3B	GUA / Val	1.172	0.416
	CCG / Val	1.051	0.460
	CAA / Gln	0.746	0.507
	CGU / Arg	0.399	0.512
3B/3C	GUA / Val	1.486	0.416
	UUA / Leu	0.526	0.388
	AUA / Ile	1.276	0.443
3C/3D	AUA / Ile	0.623	0.443

*RSCU value was calculated by the codon usage of human.

DISCUSSION

In this study, we identified similarities in synonymous codon usage between the sequences of the RNA virus HAV and those in humans, which may indicate that despite regular

mutations in the HAV genome, the synonymous codon usage of this virus is maintained similar to that of its host. This feature may help HAV to ensure successful and efficient replication through the translational machinery of human cells. Moreover, the codons recognized by the most abundant tRNA tends to be used preferentially for a given amino acid (Bennetzen and Hall, 1982; Akashi, 1994). Large-scale codon deoptimization in the poliovirus ORF has been shown to decrease viral virulence and replication (Mueller et al., 2006; Coleman et al., 2008). Thus, synonymous codon usage in HAV similar to that in human cells may be a specific strategy for HAV to ensure efficient translation for production of viral proteins.

Our examination of the regions at the cleavage sites in the polyprotein of HAV indicated that although different cleavage sites possess different motifs, at least one residue is conserved at any given cleavage site. These conserved amino acids might serve as signals recognized by 3C^{pro}, which plays a central role in HAV polyprotein cleavage. The conservation of residues at all cleavage sites provided evidence that mutations are limited in these regions, probably because the requirement for cleavage imposes significant functional and structural limitations on the polyprotein. Because of the constraints on the size of the viral genome, the abundant evolutionary information of HAV probably had to be compressed into this size-constrained genome during its evolution. Translational regulation is controlled not only by the 5' and 3' non-translational region of the HAV polyprotein (Whetter et al., 1994; Pintó et al., 2012), but also by the synonymous codon usage bias in the specific positions within the entire coding sequence. Optimal codons increase the efficiency of gene translation, whereas rare codons somewhat decrease the speed of ribosome scanning along the coding sequence (Deana et al., 1998; Rocha, 2004). Moreover, a specific synonymous codon in a specific position in the coding sequence may affect the secondary structure and biological functions of proteins (Rosano and Ceccarelli, 2009; Tuller et al., 2010). The positions with over- or under-represented codons might play an important role in regulating the rate of translation of the protein regions flanking the cleavage site in the HAV polyprotein. Furthermore, it is of interest that underrepresented codons in both HAV and human sequences, namely CUA for Leu, ACG for Thr, GUA for Val, and UCG for Ser, are preferentially used in the contexts flanking the cleavage sites of this virus. These underrepresented codons probably regulate the speed of ribosome scanning for formation of proper secondary structures of the target regions in the HAV polyprotein. This suggestion agrees with previous reports that even a single rare codon in the initiation site can control the translation efficiency of a gene (Chen and Inouye, 1994; Weygand-Durasevic and Ibba, 2010). The strong structural limitations of the cleavage sites of HAV are likely strengthened by the overusage of rare codons, whose locations may create codon context variations along the HAV mRNA, inducing a decrease in ribosome scanning speed and thus enabling proper protein folding. This may suggest that the underrepresented synonymous codons might assist new protein products in shaping the correct folding environment in the sequence contexts flanking the cleavage sites of HAV. For example, synonymous codon usage patterns control ribosome traffic and protein folding (Saunders and Deane, 2010) and the kinetics of the protein translation can affect the *in vivo* protein folding pathway, leading to a high level of protein misfolding (Komar et al., 1999; Sugiyama et al., 2005; Parmley and Huynen, 2009; Zhang and Ignatova, 2009; Weygand-Durasevic and Ibba, 2010). These observations may support our findings that the underrepresented codons with overusage in the regions flanking the HAV polyprotein cleavage likely decrease the speed of translation in these regions to ensure correct folding of proteins.

In conclusion, the observed commonalities and differences in synonymous codon usage pattern between HAV and human sequences indicate that HAV may adopt a unique translation strategy to make efficient use of the translational machinery of its human host cells. Conservation of codons at the cleavage site motifs may help HAV polyprotein to be cleaved correctly. In addition, both over- and underrepresented codons in the sequence contexts flanking the cleavage sites of HAV may play an important role in regulating the speed of translation and protein folding from these sequences in the virus.

ACKNOWLEDGMENTS

Research supported by grants from the National Natural Science Foundation of China (#31160033).

[Supplementary material](#)

REFERENCE

- Akashi H (1994). Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136: 927-935.
- Aragonès L, Bosch A and 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.
- Aragonès L, Guix S, Ribes E, Bosch A, et al. (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 and Linial M (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* 5: 311.
- Barrai I, Salvatorelli G, Mamolini E, De Lorenzi S, et al. (2008). General preadaptation of viral infectors to their hosts. *Intervirology* 51: 101-111.
- Bennetzen JL and Hall BD (1982). Codon selection in yeast. *J. Biol. Chem.* 257: 3026-3031.
- Buenz EJ and Howe CL (2006). Picornaviruses and cell death. *Trends Microbiol.* 14: 28-36.
- Chen GT and Inouye M (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* 8: 2641-2652.
- Cohen JI, Ticehurst JR, Purcell RH, Buckler-White A, et al. (1987). Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. *J. Virol.* 61: 50-59.
- Coleman JR, Papamichail D, Skiena S, Futcher B, et al. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science* 320: 1784-1787.
- Cooper DN and Youssoufian H (1988). The CpG dinucleotide and human genetic disease. *Hum. Genet.* 78: 151-155.
- De Amicis F and Marchetti S (2000). Intercodon dinucleotides affect codon choice in plant genes. *Nucleic Acids Res.* 28: 3339-3345.
- Deana A, Ehrlich R and Reiss C (1998). Silent mutations in the *Escherichia coli* ompA leader peptide region strongly affect transcription and translation *in vivo*. *Nucleic Acids Res.* 26: 4778-4782.
- Drake JW and Holland JJ (1999). Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* 96: 13910-13913.
- Kane JF (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 6: 494-500.
- Karlin S and Burge C (1995). Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet.* 11: 283-290.
- Komar AA, Lesnik T and Reiss C (1999). Synonymous codon substitutions affect ribosome traffic and protein folding during *in vitro* translation. *FEBS Lett.* 462: 387-391.
- Mueller S, Papamichail D, Coleman JR, Skiena S, et al. (2006). Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* 80: 9687-9696.
- Nakamura Y, Gojobori T and Ikemura T (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28: 292.
- Ohno H, Sakai H, Washio T and Tomita M (2001). Preferential usage of some minor codons in bacteria. *Gene* 276: 107-

115.

- Parmley JL and Huynen MA (2009). Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. *PLoS Genet.* 5: e1000548.
- Pintó RM, Aragonès L, Costafreda MI, Ribes E, et al. (2007). Codon usage and replicative strategies of hepatitis A virus. *Virus Res.* 127: 158-163.
- Pintó RM, D'Andrea L, Pérez-Rodríguez FJ, Costafreda MI, et al. (2012). Hepatitis A virus evolution and the potential emergence of new variants escaping the presently available vaccines. *Future Microbiol.* 7: 331-346.
- Rocha EP (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* 14: 2279-2286.
- Rosano GL and Ceccarelli EA (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. *Microb. Cell Fact.* 8: 41.
- Sablok G, Nayak KC, Vazquez F and Tatarinova TV (2011). Synonymous codon usage, GC(3), and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Mol. Biotechnol.* 49: 116-128.
- Sánchez G, Bosch A and Pintó RM (2003). Genome variability and capsid structural constraints of hepatitis a virus. *J. Virol.* 77: 452-459.
- Saunders R and Deane CM (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res.* 38: 6719-6728.
- Schultheiss T, Kusov YY and Gauss-Müller V (1994). Proteinase 3C of hepatitis A virus (HAV) cleaves the HAV polyprotein P2-P3 at all sites including VP1/2A and 2A/2B. *Virology* 198: 275-281.
- Schultheiss T, Sommergruber W, Kusov Y and Gauss-Müller V (1995). Cleavage specificity of purified recombinant hepatitis A virus 3C proteinase on natural substrates. *J. Virol.* 69: 1727-1733.
- Sharp PM and Li WH (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* 14: 7737-7749.
- Song H, Zhou L, Fang W, Li Y, et al. (2004). High-level expression of codon optimized foot-and-mouth disease virus complex epitopes and cholera toxin B subunit chimera in *Hansenula polymorpha*. *Biochem. Biophys. Res. Commun.* 315: 235-239.
- Sugiyama T, Gursel M, Takeshita F, Coban C, et al. (2005). CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+CD11c+ monocytes. *J. Immunol.* 174: 2273-2279.
- Tuller T, Waldman YY, Kupiec M and Ruppin E (2010). Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U. S. A.* 107: 3645-3650.
- Weygand-Durasevic I and Ibba M (2010). Cell biology. New roles for codon usage. *Science* 329: 1473-1474.
- Whetter LE, Day SP, Elroy-Stein O, Brown EA, et al. (1994). Low efficiency of the 5' nontranslated region of hepatitis A virus RNA in directing cap-independent translation in permissive monkey kidney cells. *J. Virol.* 68: 5253-5263.
- Wong EH, Smith DK, Rabadan R, Peiris M, et al. (2010). Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC Evol. Biol.* 10: 253.
- Zhang G and Ignatova Z (2009). Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS One* 4: e5036.
- Zhi N, Wan Z, Liu X, Wong S, et al. (2010). Codon optimization of human parvovirus B19 capsid genes greatly increases their expression in nonpermissive cells. *J. Virol.* 84: 13059-13062.
- Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2011). The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus. *Infect. Genet. Evol.* 11: 1815-1819.
- Zhou JH, Gao ZL, Zhang J, Chen HT, et al. (2012). Comparative [corrected] codon usage between the three main viruses in pestivirus genus and their natural susceptible livestock. *Virus Genes* 44: 475-481.
- Zhou JH, Gao ZL, Zhang J, Ding YZ, et al. (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, et al. (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.* 16C: 270-274.
- Zolotukhin S, Potter M, Hauswirth WW, Guy J, et al. (1996). A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J. Virol.* 70: 4646-4654.