Acta virologica **62**: 441 – 446, 2018 doi:10.4149/av_2018_413

The effects of nucleotide usage in key nucleotide positions +4 and -3 flanking start codon on translation levels mediated by IRES of hepatitis C virus

P. MA[1#], X.-X. MA[1#], Q.-Y. CHANG[1], L.-J. LI[1], Y.-N. WANG[2], Y.-P. FENG[1], Z.-R. MA[1*], J.-H. ZHOU[2*]

[1]Center for Biomedical Research, Northwest Minzu University, Lanzhou, 730030, P. R. China;
[2]State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046, Gansu, P. R. China

Received December 12, 2017; accepted March 20, 2018; accepted October 15, 2018

Summary. – Internal ribosomal entry site (IRES) functions as a *cis*-acting RNA element, which drives an alternative and cap-independent translation initiation pathway. Currently, there are few studies on effects of nucleotide usages at key nucleotide positions +4 and -3 flanking start codon mediated by IRES of hepatitis C virus (HCV). Herein, we focus on the effect of nucleotide usages at -3 and +4 positions mediated by HCV IRES. The nucleotide contexts flanking AUG start codon employed by HCV IRES is firstly analyzed. We found that each position in the six nucleotide positions (-4 to +6) flanking start codon of HCV has a strong tendency to select the specific nucleotide. A set of bicistronic expression vectors containing *CAT* gene, HCV IRES and *EGFP* gene were constructed, including 16 different nucleotide combinations at position -3 and +4. Each set, in which nucleotide at the -3 and +4 position has been changed into different nucleotides, included 16 types of bicistronic expression vectors. It was found that the purine nucleotide at the position -3 or +4 obviously impacts on HCV IRES-related expression, and IRES-driven translation is potentially influenced by the Kozak rule. Our results suggest that optimization of nucleotides at positions -3 and +4 is a convenient and efficient way to enhance the level of IRES-mediated translation.

Keywords: Cap-independent translation; internal ribosomal entry site; hepatitis C virus; bicistronic expression vector; translation efficiency

Introduction

The hepatitis C virus (HCV) is one of the most common blood-borne pathogens and results in acute and chronic liver diseases (Szabo *et al*., 2015). Due to an error-prone replication mechanism mediated by the viral-encoded RNA polymerase and the ability to survive mutation, HCV evolves rapidly and has extensive sequence diversity (Choo *et al*., 1991; Yuan *et al*., 2017). In 5' untranslated region (5' UTR), the internal ribosomal entry site (IRES) of HCV is the most highly conserved region among various genotypes, its IRES plays an important role in the viral cycle and provides a potent target for antiviral agents (Fraser and Doudna, 2007). As for translation initiation mechanisms, there are two major mechanisms in eukaryotic translation systems, namely 5' cap-dependent translation initiation mechanism and 5' cap-independent mechanism mediated by internal ribosome entry site (IRES) element (Mailliot and Martin, 2017). It has been accepted that initiation of translation of HCV RNA starts by a cap-independent mechanism mediated by IRES. Compared with the mechanism of translation initiation mediated by 5' cap structure [methyl-7-G(5')pppG(5')N] (Green and Noller, 1997), the highly ordered structures within IRES elements are absolutely required for IRES activity (Balvay *et al*., 2009). By the use of a toe-print assay to determine the position of the ribosome on the mRNA, the HCV IRES was found to recruit the small ribosomal

*Corresponding author. E-mail: zhoujianhua03@caas.cn; phone: +86-931-2938311. #The two authors contributed equally to this work.

Abbreviations: CAT = chloramphenicol acetyl transferase; EGFP = enhanced green fluorescent protein; HCV = hepatitis C virus; IRES = internal ribosomal entry site

subunit directly to the ribosome, scanning to the initiation codon would not be required to initiate protein synthesis. Furthermore, the introduction of an initiation codon that is 7 nucleotides upstream, or 8 nucleotides downstream, of the correct initiation codon is not recognized by the ribosome, indicating that the structure of the HCV IRES is highly specialized in placing the correct initiation codon close to the 40S ribosomal peptidyl (P) site. As for the secondary structure of domain IV of HCV IRES (Fraser and Doudna, 2007), this domain includes a loop, which is composed of 7 nucleotides, namely the (-3)AUC AUG (+4)A motif in positions -3 to +4. As for translation initiation mediated by cap-dependent mechanism, nucleotide context surrounding start codon influences the situation of ribosome scanning (Kozak, 1986; Kozak, 1987). Mutations occurring at position -3 and +4 of this nucleotide context can weaken the abilities of ribosomes to recognize the corresponding start codon (Kozak, 1997; Kozak, 2003). The mechanism of ribosomal recruitment onto the IRES element of HCV is significantly different from that mediated by 5' cap-structure. Without the participation of eIF4F, eIF4A and eIF4B, the 48S initiation complex can be assembled onto the HCV IRES element (Kieft *et al.*, 2001; Otto *et al.*, 2002; Laletina *et al.*, 2006). During the course of the positioning of the IRES element of HCV onto the surface of the 40S ribosomal subunit, the specific nucleotide sequence structure of the IRES element can induce a conformational change of 40S ribosome, resulting in positioning the start codon into the P site of the ribosome without ribosomal scanning (Pestova *et al.*, 1998; Spahn *et al.*, 2001). Here, we are interested in whether or not the specific nucleotide usage in positions -3 and +4 flanking the start codon would take part in mediating the translation initiation efficiencies under the HCV IRES element. Currently, there have been many studies on the role of functional and structural diversity of the HCV IRES element in mediating the translation initiation, however, there are a few reports about the specific nucleotide positions flanking the start codon under the control of HCV IRES element. In this study, we focused on the nucleotide usage at positions -3 and +4 surrounding the start codon for influencing the downstream gene expression level in order to provide insights into the translation initiation mediated by the HCV IRES element.

Materials and Methods

HCV genomes data. To reflect the genetic diversity for HCV genome, the 273 genomes of HCV, which were collected from different countries at different times, were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/) (listed in Table S1).

Analysis of nucleotide usage at translation initiation region of HCV genome. To analyze the features of nucleotide usage at positions-3 & +4 for HCV genome, we employed a simple method to estimate the magnitudes of nucleotide usage at the translation initiation region (positions -3, -2, -1 and +4, +5 and +6), namely $R = n_i / N_{opt}$. In detail, R_i represents the usage magnitude of the nucleotide *i* at the target position of the 273 HCV sequences, n_i is the frequency of the interesting nucleotide *i* at the target position, N_{opt} is the frequency of the optimal nucleotide at this position. The formula for R_i was calculated for estimating the usage magnitudes of the four types of nucleotides at each position of translation initiation region.

Plasmids construction. To estimate the effects of each type of nucleotides at the positions -3 & +4 on the translation initiation efficiencies mediated by the HCV IRES element, we downloaded the HCV IRES sequence (GenBank Acc. No. AB016785) from GenBank of NCBI website and this IRES sequence was artificially synthesized by the GENEWIZ (China) Co., Ltd. Bicistronic reporter constructs were engineered to contain the IRES element. The upstream chloramphenicol acetyl transferase (CAT) and downstream enhanced green fluorescent protein (EGFP) genes monitor cap-dependent and IRES-dependent translation, respectively. It was noted that the target IRES sequence was fused in frame of CAT coding sequence. During constructing the bicistronic reporter constructs containing the specific IRES, all performances were carried out via standard techniques, following Molecular Cloning (A Laboratory Manual, 3rd ed. Cold Spring Harbor). First step was that acquisitions of CAT coding sequence, IRES sequence and EGFP coding sequence were achieved by PCR method with the specific forward & reverse PCR primers (Table S2). In detail, CAT gene was amplified from the plasmid pcDNA3.1(+)/CAT (Invitrogen), EGFP gene was amplified from the plasmid pEGFP-N1 (CLONTECH Laboratories, Inc.), and the HCV IRES element was obtained from plasmid pUC57-IRES(III) (Invitrogen). Second step is that the CAT gene and the IRES sequence were amplified via fusion PCR with a set of primers to obtain the cloned fragments CAT-IRES (Table S3). The fusion PCR amplifications involved 35 cycles of 94°C for 10 s, 58°C for 15 s, and 68°C for 70 s, followed by a final elongation step at 72°C for 10 min. Third step is that the fragment of CAT-IRES and EGFP coding sequence were carried out via fusion PCR with a set of primers to obtain the cloned fragments CAT-IRES-EGFP. The program of fusion PCR amplifications was the same as that of the fusion PCR for the cloned fragment of CAT-IRES (Table S4). The resulting cloned fragments were processed by digestion with B*am*HI and X*ba*I and ligated into similarly digested pcDNA3.1 (+). There were 16 bicistronic reporter constructs, which contain the specific translation initiation regions. By sequencing assay of the two resulting plasmids, the 16 resulting bicistronic reporter constructs were shown to contain the correct fragments (Table 1).

Transient expression assay. Because the 16 bicistronic reporter plasmids contained different nucleotide usage at the position +4 in the translation initiation region, the different nucleotide usage can result in different codon usages, which were located downstream of the start codon. The factor of tRNA abundance pairing to the specific codon needed to be analyzed whether or not it would work in influencing EGFP gene expression. As for the cell culture using

Table 1. The information about the nucleotide context of translation initiation regions in the 16 bicistronic reporter constructs

No.	Name of the bicistronic reporter construct	The nucleotide context
1	pCAT-IRES$_1$-EGFP	(-3)**T**CC ATG (+4)**G**TG
2	pCAT-IRES$_2$-EGFP	(-3)**T**CC ATG (+4)**C**TG
3	pCAT-IRES$_3$-EGFP	(-3)**T**CC ATG (+4)**T**TG
4	pCAT-IRES$_4$-EGFP	(-3)**T**CC ATG (+4)**A**TG
5	pCAT-IRES$_5$-EGFP	(-3)**C**CC ATG (+4)**G**TG
6	pCAT-IRES$_6$-EGFP	(-3)**C**CC ATG (+4)**C**TG
7	pCAT-IRES$_7$-EGFP	(-3)**C**CC ATG (+4)**T**TG
8	pCAT-IRES$_8$-EGFP	(-3)**C**CC ATG (+4)**A**TG
9	pCAT-IRES$_9$-EGFP	(-3)**A**CC ATG (+4)**G**TG
10	pCAT-IRES$_{10}$-EGFP	(-3)**A**CC ATG (+4)**C**TG
11	pCAT-IRES$_{11}$-EGFP	(-3)**A**CC ATG (+4)**T**TG
12	pCAT-IRES$_{12}$-EGFP	(-3)**A**CC ATG (+4)**A**TG
13	pCAT-IRES$_{13}$-EGFP	(-3)**G**CC ATG (+4)**G**TG
14	pCAT-IRES$_{14}$-EGFP	(-3)**G**CC ATG (+4)**C**TG
15	pCAT-IRES$_{15}$-EGFP	(-3)**G**CC ATG (+4)**T**TG
16	pCAT-IRES$_{16}$-EGFP	(-3)**G**CC ATG (+4)**A**TG

as the target cell of transient transfection, we used cell line CHO-K1 (Chinese hamster ovary), because the tRNA copy numbers of CHO-K1 had been reported in the Genomic tRNA Database (http://gtrnadb.ucsc.edu/#eukarya). CHO-K1 cells in 6-well plate were transfected one day after plating, when the monolayers reached to about 80% confluence. Following the protocol provided with Lipofectamine™ 2000 Reagent (Invitrogen), each plate was transfected with 500 µl of Opti-MEM™ Medium mixture including 4 µl of Lipofectamine™ 2000 and 4 µg of purified plasmid DNA. Twenty-four hours after transfection, the cells were collected. Soluble extracts of CHO-K1 cells were prepared by lysing the cells in 120 mmol/l NaCl, 50 mmol/l Tris-HCl (pH 7.8), and 0.5% of Nonidet P-40, followed by centrifugation at 12,000 rpm for 5 min in a 1.5-ml centrifuge tube. These products were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; 10% gels) and detected by immunoblotting with a rabbit anti-CAT antibody (Sigma) and murine anti-EGFP antibody (Sigma) with horseradish peroxidase (HRP)-labeled anti-species secondary antibodies (Sigma) and enhanced chemiluminescence (ECL; Thermo).

Data analysis. To analyze the translation magnitudes mediated by different nucleotides at the positions -3 and +4, the products translated by CAT gene and EGFP gene were tested and confirmed. The translation ratio (R) was calculated by the formula R = EGFP/CAT. The results were analyzed by the ANOVA test performed by SPSS 11.5 software and were considered significantly different when $P < 0.05$.

Results

Nucleotide usage at the positions -3 and +4 surrounding start codon

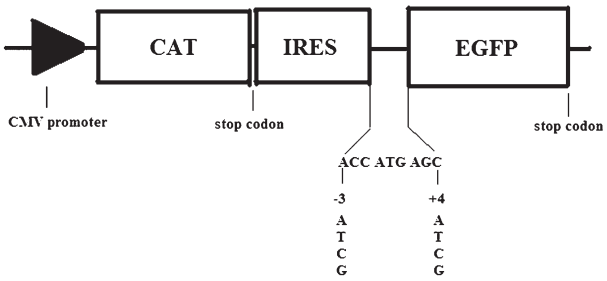
By analyzing the two target positions surrounding the start codon, we found the bias of nucleotide usage existed in translation initiation region of HCV genome. As for the target positions of this study, adenine had strong tendencies to exist at the position -3 and +4 (Table 2). This result showed that nucleotide usage at the two positions is under strong natural selection rather than mutation pressure, even though HCV as an RNA virus has high mutation rates.

The effect of the key nucleotide context flanking the start codons on gene transcription

To explore the effect of -3 and +4 mutations on IRES mediated translation, we developed bicistronic expression vector, in which single transcripts contained the CAT and EGFP genes under the control of human cytomegalovirus (CMV) promoter. As shown in Fig. 1, the IRES was inserted into the junction of these two genes, and single or double mutations of positions -3 and +4 flanking the start codon used by IRES for the translation initiation of EGFP were introduced into IRES-based constructs, resulting in 16 constructs with different combinations for each IRES. As for these two reporter genes, CAT and EGFP, both were detected in IRES-based constructs, from which CAT was translated by a cap-independent manner and EGFP translation as mediated by HCV IRES element. Real-time RT-PCR was performed at 24 hours post-transfection in raw cell extracts

Table 2. Frequencies and relative adaptiveness (R_i) of the nucleotide surrounding the start codon

Position	F_A	F_U	F_C	F_G	R_A	R_U	R_C	R_G
-3	235	1	1	0	1.00	0.004255	0.004255	0
-2	13	55	169	0	0.076923	0.325444	1.00	0
-1	1	1	23	0	0.043478	0.043478	1.00	0
AUG	-------	-------	-------	--------	-------	--------	--------	--------
+4	235	0	0	2	1.00	0.00	0.00	0.008511
+5	1	0	1	235	0.004255	0.00	0.004255	1.00
+6	1	0	236	0	0.004237	0.00	1.00	0.00



pCAT-IRES-EGFP

Fig. 1

The organization of bicistronic expression vector (pCAT-IRES-EGFP)

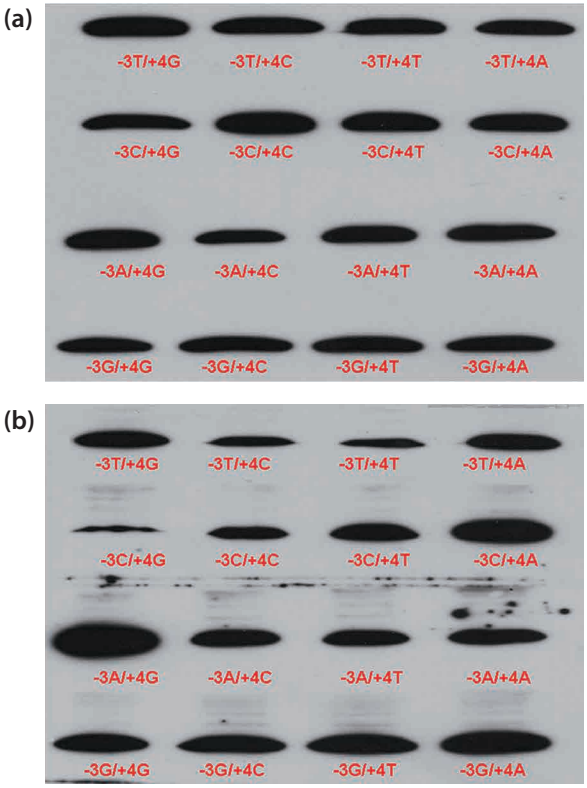


Fig. 2

Western blot assay for the detection of the protein product of CAT gene and EGFP gene translated by 16 different sets of pCAT-IRES-EGFP

(a) CAT protein products, **(b)** EGFP protein products.

transfected with different bicistronic reporter constructs. The results showed similar levels of reporter RNAs in each reaction (Fig. S1), suggesting that the changes of nucleotides at the positions -3 and +4 did not affect the transcription levels of reporter genes.

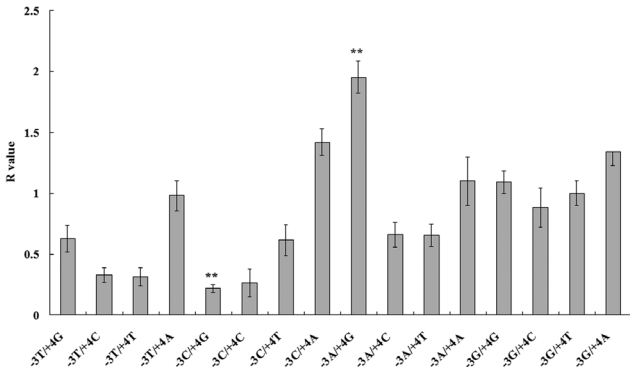


Fig. 3

Quantification of CAT and EGFP protein expression translated by 16 different sets of pCAT-IRES-EGFP vectors caused by the changes of nucleotide usages at positions -3 and +4 surrounding the start codon

The effect of the key nucleotide context flanking the start codons on gene expression

Twenty-four hours after transient transfection, the 16 raw cell extracts transfected by the specific bicistronic reporter construct were collected. Through western blotting assay we showed that the two reporter genes can be translated in CHO cells (Fig. 2). Furthermore, while CAT gene expression levels were similar, the expression levels of EGFP gene mediated by different nucleotides at the positions -3 and +4 were different. As for the translation ratios affected by different nucleotides at the positions of interest, the relative expression ratio of EGFP to CAT was calculated to represent the relative translation efficiency of IRES with specific nucleotide context around AUG start codon. In comparison, the relative translation efficiency (R value) driven by HCV IRES (-3A, +4G) was highest ($p<0.05$), however, the relative translation efficiency (R value) mediated by HCV IRES (-3C, +4G) was lowest ($P <0.05$) (Fig. 3). In addition, as shown in Fig. 3, when the position +4 was nucleotide A, the expression of EGFP mediated by HCV IRES could be improved regardless of the nucleotide types in the position -3; and when the position -3 was nucleotide G, the expression of EGFP also could be enhanced, regardless of the nucleotide type in the position +4. These results strongly implied that the nucleotide related to purine at the position -3 or +4 impacted the expression of gene mediated by HCV IRES.

Discussion

From the previous detailed mutational analyses, the Kozak sequence (-3A/GCCATGG+4) has been defined as the most favorable context for translation initiation in eukaryotes mediated by 5'cap-dependent mechanism under normal

conditions (Hendrickx *et al.*, 2014; Mohan *et al.*, 2014; Ferretti *et al.*, 2017). Of all the nucleotides around the AUG start codon, the nucleotides at positions -3 and +4 have dominant effects on translation efficiency (Kozak, 2003). Herein, this study shows that the positions -3 and +4 surrounding the start codon play important roles in gene expression mediated by HCV IRES. Compared with the gene expression mediated by HCV IRES (-3A, +4A), the EGFP gene expression mediated by IRES (-3A, +4G), which is similar with that of the key positions of Kozak sequence, is higher. To some degree, the tRNA abundances are always impacting the translation efficiency of the specific gene, because translation speed and ribosomal density are encoded by coding sequences and the tRNA abundance pool (Tuller *et al.*, 2010; Sun *et al.*, 2018). According to tRNA copy numbers for the tRNA abundance pool in CHO cell line (Genomic tRNA Database), when the position +4 mutates different nucleotide types in bicistronic reporter plasmids, the tRNA abundances for the second codon (+4XTG+6, X means any nucleotide, tRNA copy number for **G**TG is 9, tRNA copy number for **C**TG is 10, tRNA copy number for TTG is 4, tRNA copy number for **A**TG is 16) following the start codon have no obvious effect on mediating the EGFP gene expression levels. This result implies that nucleotide context at +4 nucleotide position might influence the local structural formation rather than tRNA abundances, leading to the changes of gene expression mediated by HCV IRES (Jaafar *et al.*, 2016; Bugaud *et al.*, 2017; Ma *et al.*, 2018; Mengardi *et al.*, 2017; Ross *et al.*, 2017; Zhu *et al.*, 2017). Of note, the results show that the positions -3A and +4G had a better performance in improving the translation efficiency mediated by HCV IRES than the preferred nucleotide usage bias (-3A, +4A) in the positions -3 & +4 of the translation initiation region of HCV, implying that the Kozak-like nucleotide context (-3A, +4G) probably impacts the ribosome binding to the translation initiation region of HCV. The HCV IRES has a capability to direct the initiator tRNA to the ribosomal P site during internal initiation and this IRES, which depends on modular architecture, plays important role in the gene expression by cap-independent mechanisms (Yamamoto *et al.*, 2015).

The usage of IRES in bicistronic expression plasmids enables the expression of two genes controlled by one promoter in the same cells, which are important tools in today's cell biology (Fajardo *et al.*, 2012; Ma *et al.*, 2016). In order to obtain ratio-controlled or improved IRES-mediated protein expression, many methods were developed, for example, removal of unnecessary AUG start codon, using multiple IRES elements (Bouabe *et al.*, 2008; Koh *et al.*, 2013). On the basis of our data, it is clear that nucleotide mutations at position -3 and +4 flanking start codon could be an efficient and simple way to achieve maximal protein translation when viral IRES are used for heterologous co-expression. Since the IRES element exhibits cell- and tissue-specificity, our study broadens the choices of IRES and it can be recommended to select the IRES to be used for co-expression according to the cell type or tissue, rather than using encephalomyocarditis virus (EMCV) IRES, which is commonly proposed in all commercial IRES-based vectors.

Acknowledgments. This study was supported by the National Natural Science foundation of China (No. 31302100).

Supplementary information is available in the online version of the paper.

References

Balvay L, Soto Rifo R, Ricci EP, Decimo D, Ohlmann T (2009): Structural and functional diversity of viral IRESes. Biochim. Biophys. Acta 1789, 542–557. https://doi.org/10.1016/j.bbagrm.2009.07.005

Bouabe H, Fassler R, Heesemann J (2008): Improvement of reporter activity by IRES-mediated polycistronic reporter system. Nucleic. Acids. Res. 36, e28. https://doi.org/10.1093/nar/gkm1119

Bugaud O, Barbier N, Chommy H, Fiszma N, Le Gall A, Dulin D, Saguy M, Westbrook N, Perronet K, Namy O (2017): Kinetics of CrPV and HCV IRES-mediated eukaryotic translation using single-molecule fluorescence microscopy. RNA 23, 1626–1635. https://doi.org/10.1261/rna.061523.117

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby R, Barr PJ, et al (1991): Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA 88, 2451–2455. https://doi.org/10.1073/pnas.88.6.2451

Fajardo T, Jr, Rosas MF, Sobrino F, Martinez-Salas E (2012): Exploring IRES region accessibility by interference of foot-and-mouth disease virus infectivity. Plos One 7, e41382. https://doi.org/10.1371/journal.pone.0041382

Ferretti MB, Ghalei H, Ward EA, Potts EL, Karbstein K (2017): Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. Nat. Struct. Mol. Biol. 24, 700–707. https://doi.org/10.1038/nsmb.3442

Fraser CS, Doudna JA (2007): Structural and mechanistic insights into hepatitis C viral translation initiation. Nat. Rev. Microbiol. 5, 29–38. https://doi.org/10.1038/nrmicro1558

Green R, Noller HF (1997): Ribosomes and translation. Annu. Rev. Biochem. 66, 679–716. https://doi.org/10.1146/annurev.biochem.66.1.679

Hendrickx G, Boudin E, Fijalkowski I, Nielsen TL, Andersen M, Brixen K, Van Hul W (2014): Variation in the Kozak sequence of WNT16 results in an increased translation and is associated with osteoporosis related parameters. Bone 59, 57–65. https://doi.org/10.1016/j.bone.2013.10.022

Jaafar ZA, Oguro A, Nakamura Y, Kieft JS (2016): Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. eLife 5. https://doi.org/10.7554/eLife.21198

Kieft JS, Zhou K, Jubin R, Doudna JA (2001): Mechanism of ribosome recruitment by hepatitis C IRES RNA. RNA 7, 194–206. https://doi.org/10.1017/S1355838201001790

Koh EY, Ho SC, Mariati Song Z, Bi X, Bardor M, Yang Y (2013): An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells. Plos One 8, e82100. https://doi.org/10.1371/journal.pone.0082100

Kozak M (1986): Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283–292. https://doi.org/10.1016/0092-8674(86)90762-2

Kozak M (1987): An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic. Acids. Res. 15, 8125–8148. https://doi.org/10.1093/nar/15.20.8125

Kozak M (1997): Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J. 16, 2482–2492. https://doi.org/10.1093/emboj/16.9.2482

Kozak M (2003): Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation. Gene 318, 1–23. https://doi.org/10.1016/S0378-1119(03)00774-1

Laletina E, Graifer D, Malygin A, Ivanov A, Shatsky I, Karpova G (2006): Proteins surrounding hairpin IIIe of the hepatitis C virus internal ribosome entry site on the human 40S ribosomal subunit. Nucleic. Acids. Res. 34, 2027–2036. https://doi.org/10.1093/nar/gkl155

Ma XX, Feng YP, Gu YX, Zhou JH, Ma ZR (2016): Effect of the nucleotides surrounding the start codon on the translation of foot-and-mouth disease virus RNA. Acta Virol. 60, 151–155. https://doi.org/10.4149/av_2016_02_151

Ma XX, Ma Z, Pan Q (2018): The challenges of long-term transcriptional gene silencing by RNA virus. Trends Biochem. Sci. 43, 649–650. https://doi.org/10.1016/j.tibs.2018.06.010

Mailliot J, Martin F (2017): Viral internal ribosomal entry sites: four classes for one goal. Wiley Interdiscip. Rev. RNA. doi: 10.1002/wrna.1458

Mengardi C, Limousin T, Ricci EP, Soto-Rifo R, Decimo D, Ohlmann T (2017): microRNAs stimulate translation initiation mediated by HCV-like IRESes. Nucleic. Acids. Res. 45, 4810–4824. https://doi.org/10.1093/nar/gkw1345

Mohan RA, van Engelen K, Stefanovic S, Barnett P, Ilgun A, Baars MJ, Bouma BJ, Mulder BJ, Christoffels VM, Postma AV (2014): A mutation in the Kozak sequence of GATA4 hampers translation in a family with atrial septal defects. Am. J. Med. Genet. A. 164A, 2732–2738. https://doi.org/10.1002/ajmg.a.36703

Otto GA, Lukavsky P J, Lancaster AM, Sarnow P, Puglisi JD (2002): Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction. RNA 8, 913–923. https://doi.org/10.1017/S1355838202022057

Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU (1998): A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. Genes Dev. 12, 67–83. https://doi.org/10.1101/gad.12.1.67

Ross MJ, Fidai I, Cowan JA (2017): Analysis of Structure-Activity Relationships Based on the Hepatitis C Virus SLIIb Internal Ribosomal Entry Sequence RNA-Targeting GGHYR-FKCu Complex. Chembiochem. 18, 1743–1754. https://doi.org/10.1002/cbic.201700228

Spahn CM, Kieft JS, Grassucci RA, Penczek PA, Zhou K, Doudna JA, Frank J (2001): Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. Science 291, 1959–1962. https://doi.org/10.1126/science.1058409

Sun C, Yang F, Zhang Y, Chu J, Wang Y, Zhang Y, Li J, Li Y, Fan R, Li W, Huang X, Wu H, Fu Z, Jiang Z, Yin Y (2018): tRNA-derived fragments as novel predictive biomarkers for trastuzumab-resistant breast cancer. Cell. Physiol. Biochem. 49, 419–431. https://doi.org/10.1159/000492977

Szabo SM, Samp JC, Walker DR, Lane S, Cline SK, Gooch KL, Jimenez-Mendez R, Levy AR (2015): Liver-specific case fatality due to chronic hepatitis C virus infection: a systematic review. Ann. Hepatol. 14, 618–630.

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y (2010): An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141, 344–354. https://doi.org/10.1016/j.cell.2010.03.031

Yamamoto H, Collier M, Loerke J, Ismer J, Schmidt A, Hilal T, Sprink T, Yamamoto K, Mielke T, Burger J, Shaikh TR, Dabrowski M, Hildebrand PW, Scheerer P, Spahn CM (2015): Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. EMBO J. 34, 3042–3058. https://doi.org/10.15252/embj.201592469

Yuan G, Liu J, Hu C, Huang H, Qi M, Wu T, Liang W, Li YP, Zhang YY, Zhou Y (2017): Genotype distribution and Molecular Epidemiology of hepatitis C virus in Guangzhou, China.: Predominance of genotype 1b and increasing incidence of genotype 6a. Cell. Physiol. Biochem. 43, 775–787. https://doi.org/10.1159/000481561

Zhu Y, Huang P, Yang N, Liu R, Liu X, Dai H, Zhang L, Song F, Sun C (2017): Establishment and Application of a High Throughput Screening System Targeting the Interaction between HCV Internal Ribosome Entry Site and Human Eukaryotic Translation Initiation Factor 3. Front. Microbiol. 8, 977. https://doi.org/10.3389/fmicb.2017.00977

Supplementary information

The effects of nucleotide usage in key nucleotide positions +4 and -3 flanking start codon on translation levels mediated by IRES of hepatitis C virus

P. MA[1#], X.-X. MA[1#], Q.-Y. CHANG[1], L.-J. LI[1], Y.-N. WANG[2], Y.-P. FENG[1], Z.-R. MA[1*], J.-H. ZHOU[2*]

[1]Center for Biomedical Research, Northwest Minzu University, Lanzhou, 730030, P. R. China;
[2]State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046, Gansu, P. R. China

Received December 12, 2017; accepted March 20, 2018; October 15, 2018

Table S1. Sequence data of HCV

Acc. No.	Subgenotypes	Acc. No.	Subgenotypes	Acc. No.	Subgenotypes
NC_004102	1a	EU781795	1a	EU155329	1b
AF011753	1a	EU781794	1a	EU155328	1b
AF011752	1a	EU781789	1a	EU155327	1b
AF011751	1a	EU781787	1a	EU155326	1b
EU862839	1a	EU781786	1a	EU155324	1b
EU255989	1a	EU781790	1a	EU482888	1b
EU255984	1a	EU781783	1a	EU155330	1b
EU255983	1a	EU781780	1a	EU155332	1b
EU255982	1a	EU781777	1a	EU256085	1b
EU255981	1a	EU781774	1a	EU256084	1b
EU255980	1a	EU781768	1a	EU256045	1b
EU255965	1a	EU781760	1a	EU256000	1b
EU255963	1a	EU256072	1a	EU256066	1b
EU155341	1a	EU256051	1a	EU155382	1b
EU255975	1a	EU256041	1a	EU155357	1b
EU255988	1a	EU256033	1a	EU155356	1b
EU155343	1a	EU256057	1a	EU155300	1b
EU155338	1a	EU256049	1a	EU482881	1b
EU255967	1a	EU256048	1a	EU256083	1b
EU255966	1a	EU255934	1a	EU256082	1b
EU781802	1a	FJ205868	1a	EU256079	1b
EU781797	1a	FJ205867	1a	EU256078	1b
EU781781	1a	EU155238	1a	EU256077	1b
EU781779	1a	EU155333	1b	EU256059	1b
EU781776	1a	EU482849	1b	EU155374	1b
EU781799	1a	EU155337	1b	FJ390398	1b
EU781798	1a	EU155334	1b	FJ390397	1b
EU781796	1a	EU155331	1b	AB249644	1b
EU781804	1a	EU155336	1b	AM910652	1g
EU781800	1a	EU155335	1b	AB047644	2a

Acc. No.	Subgenotypes	Acc. No.	Subgenotypes	Acc. No.	Subgenotypes
AB047643	2a	EU687194	1a	AY232740	2b
AB047642	2a	AB520610	1a	AY232739	2b
AB047641	2a	EU255945	1a	AY232738	2b
AB047640	2a	EU155346	1a	D50409	2c
AY232731	2b	EU482873	1a	AB031663	2k
AY232730	2b	D90208	1b	D17763	3a
AB559564	2b	D11355	1b	D28917	3a
HM777359	2j	D10934	1b	AF046866	3a
FN666428	2q	EF638081	1b	X76918	3a
GQ275355	3a	AY460204	1b	D49374	3b
FJ407092	3i	AF165064	1b	D63821	3k
FJ025854	4b	HCU89019	1b	Y11604	4a
EU392172	4d	AF483269	1b	DQ418789	4a
EU392169	4f	EU857431	1b	DQ418788	4a
EU392173	4k	L02836	1b	DQ418787	4a
EU392171	4k	GU451224	1b	DQ418784	4a
NC_009826	5a	GU451222	1b	DQ418783	4a
DQ278892	6	GU451221	1b	DQ418782	4a
DQ278894	6n	GU451220	1b	DQ516084	4a
EF424625	6q	GU451219	1b	Y13184	5a
EU408328	6r	GU451218	1b	AF064490	5a
EU408329	6s	AF176573	1b	AY859526	6a
EF632071	6t	U01214	1b	Y12083	6a
EU246940	6u	AF333324	1b	DQ480518	6a
EU643834	6w	AB426117	1b	DQ480517	6a
M62321	1a	D63857	1b	DQ480516	6a
M67463	1a	EF032894	1b	DQ480513	6a
AF009606	1a	EF032893	1b	DQ480512	6a
EF032886	1a	EF032892	1b	DQ480524	6a
EF407457	1a	AY051292	1c	DQ480523	6a
EF407456	1a	D14853	1c	DQ480520	6a
EF407455	1a	AY651061	1c	D84262	6b
EF407454	1a	D00944	2a	D84263	6d
EF407453	1a	AB047639	2a	DQ835764	6f
EF407452	1a	NC009823	2a	DQ835760	6f
EF407451	1a	AY746460	2a	D63822	6g
EF407450	1a	AF238485	2a	DQ314806	6g
EF407449	1a	AF238484	2a	D84265	6h
EU155295	1a	AF238483	2a	D84264	6k
EU155293	1a	AF238482	2a	DQ278893	6k
EU155292	1a	AF238481	2a	DQ278891	6k
EU155291	1a	AF169005	2a	EU408332	6u
EU155290	1a	D10988	2b	EU408331	6u
EU155289	1a	AB030907	2b	EU408330	6u
EU155288	1a	AY232745	2b	EU798760	6v
EF032883	1a	AY232744	2b	FJ435090	6v
EU256074	1a	AY232743	2b	EU798761	6v
EU155347	1a	AY232742	2b	EU158186	6v
EU256106	1a	AY232741	2b	EF108306	7a

Table S2. Oligonucleotides for PCR obtaining CAT, EGFP and IRES sequences

Primer	Sequence (5'→3')
F-CAT	ATGGAGAAAAAAATCACTGGATATA
R-CAT	TTACGCCCCGCCCTGCCACTCATCG
F-EGFP	ATGGAAGTGAGCAAGGGCGAGGAGCTGTTCAC
R-EGFP	TTACTTGTACAGCTCGTCCATGCCGAGAG
F-IRES	ATGAGTGGCAGGGCGGGGCGTAACCTGTGAGGAACTACTGTCTTCA
R-IRES	AGCTCCTCGCCCTTGCTCACCATGGTGCACGGTCTACGAGACCTCC

Table S3. Oligonucleotides used for fusion PCR obtaining the fragment of CAT-IRES

Primer	Sequence (5'→3')
F-CAT	5'-CGGGATCCATGGAGAAAAAAATCACTGG-3'
R-CAT	5'-TTACGCCCCGCCCTGCCACTC-3'
F-IRES	5'-TGGCAGGGCGGGGCGTAACCTGTGAGGAACTACTGTCTTC-3'
R-IRES	5'-GCACGGTCTACGAGACCTCC-3'

Note: The primer F-CAT contains *Bam*HI site. The primer F-IRES contains the 3'end of CAT coding sequence and the 5'end of IRES sequence.

Table S4. Oligonucleotides used for fusion PCR obtaining the fragment of CAT-IRES-EGFP

Primer	Sequence (5'→3')
F-CAT	5'-CGGGATCCATGGAGAAAAAAATCACTGG-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCTCCatggtgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCTCCatgctgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCTCCatgttgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCTCCatgatgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCCCCatggtgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCCCCatgctgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCCCCatgttgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCCCCatgatgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCACCatggtgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCACCatgctgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCACCatgttgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCACCatgatgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCGCCatggtgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCGCCatgctgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCGCCatgttgagcaagggcgaggagctgttcac-3'
F-EGFP*	5'-AGGTCTCGTAGACCGTGCGCCatgatgagcaagggcgaggagctgttcac-3'
R-EGFP	5'-GCTCTAGATTACTTGTACAGCTCGTCCA-3'

Note: The primer F-CAT contains *Bam*HI site. The 16 primers F-EGFP* are composed of the 3' end of IRES element and EGFP coding sequence. Each F-EGFP primer contains the specific nucleotide usage at the positions -3 and +4. The positions -3 and +4 have been marked in yellow color.

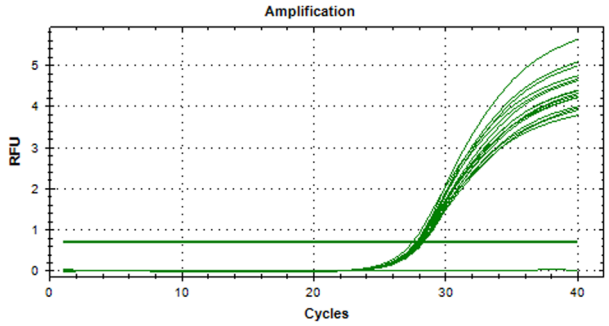


Fig. S1

Amplification plot for reporter RNAs