

脑心肌炎病毒多聚蛋白编码区起始与终止区的密码子使用模式

张瑜^{1,2}, 张峰¹, 王树瑜^{1,2}, 孙晓林^{1*}

(1. 甘肃农业大学动物医学院, 甘肃兰州 730070; 2. 中牧实业股份有限公司兰州生物药厂, 甘肃兰州 730046)

摘要 文中利用特定区域密码子使用偏嗜性对数公式分析了脑心肌炎病毒(Encephalomyocarditis virus, EMCV)多聚蛋白编码基因的翻译起始区域与翻译终止区域的密码子使用模式。结果表明,EMCV多聚蛋白编码区内翻译起始序列区的同义密码子使用偏嗜性的正负偏差总体是平衡的,但几种在整体多聚蛋白编码区域内使用为负偏嗜的稀有密码子却倾向于出现于编码起始区和终止区。这一有趣的现象,可利用“稀有密码子调控假说”来解释为EMCV开放阅读框两端的稀有密码子偏好性使用对整体阅读框的表达具有负调控作用。该研究说明“稀有密码子调控假说”不仅适用于细菌,而且也适用于一些RNA病毒基因组。

关键词 脑心肌炎病毒;翻译起始区;翻译终止区;密码子使用模式

中图分类号 S852.65 **文献标识码** A **文章编号** 0517-6611(2013)22-09202-03

The Codon Usage Pattern for Translation Initiation and Termination Regions in the Whole Coding Sequence of Encephalomyocarditis Virus

ZHANG Yu et al (College of Veterinary Medicine, Gansu Agricultural University, Lanzhou, Gansu 730070)

Abstract The codon usage pattern for the translational initiation and termination regions of encephalomyocarditis virus (EMCV) was particularly analyzed. The results indicate that the codon usage pattern with low propensity at some sites in the translation initiation and termination regions is preferentially existed, and the interesting phenomenon may imply a negative effect on gene expression, which can be explained by the “minor codon modulator hypothesis”. Obviously, some information about the pattern of codon usage in the two target regions flanking the open reading frame of HCV may indicate that the codons with low-usage may have a more obvious effect on translation process. The proposed results imply that the “minor codon modulator hypothesis” may be applied to both bacteria and some RNA viruses (e. g. EMCV).

Key words Encephalomyocarditis virus; Translation initiation region; Translation termination region; Codon usage pattern

脑心肌炎病毒(Encephalomyocarditis virus, EMCV)是一种重要的人畜共患病病原,会引起猪和某些哺乳动物、啮齿动物乃至灵长类动物的一种以脑炎、心肌炎或心肌周围炎为主要特征的急性传染病^[1]。在自然条件下,EMCV感染最广泛和最严重的动物是猪,感染后可造成突然死亡和实质器官的广泛病理损伤,母猪感染还可造成其繁殖障碍。EMCV是一种RNA病毒,属小RNA病毒科心肌炎病毒属。EMCV病毒粒子含1条单股正链RNA(长约7.8 kb),病毒粒子呈正二十面体对称,无囊膜,分子质量为 2.6×10^3 kD。EMCV的结构蛋白包括VP1、VP2、VP3和VP4,非结构蛋白有2A、2B、2C、3A、3B、3C^{pro}和3D^{pol}(依赖RNA的RNA聚合酶)^[2]。值得注意的是,在VP1的上游连接有1个无水解活性的前导蛋白,其可被3C蛋白酶水解游离出多聚蛋白。

从EMCV基因组表达出的所有病毒蛋白的顺序来看,前导蛋白和RNA依赖RNA聚合酶分别位于此病毒整体编码序列的N'-端和C'-端。基于核苷酸和氨基酸的同源性分析表明,EMCV多聚蛋白结构基因编码区域的保守性低于非结构蛋白,而3D蛋白的编码序列在所有非结构蛋白中是最保守的,这种开放阅读框5'端变异大于3'端的遗传学现象对于研究同义密码子使用模式在EMCV基因中的存在形式是十分重要的。研究EMCV开放阅读框两极的核苷酸序列组成性变化将有助于研究人员对这种病毒遗传进化特点进行深入了解,从而丰富小RNA病毒家族遗传进化的

内容^[3]。

已有研究表明,外源基因的翻译起始序列和翻译终止序列的密码子使用模式对靶基因在宿主细胞的表达调控过程起着十分重要的作用^[4]。而关于EMCV开放阅读框两端翻译起始和终止区域的密码子使用模式却未见任何报道,该研究基于EMCV整体开放阅读框的密码子使用模式来测评翻译起始和终止序列特定区域密码子使用的偏嗜性,从而为EMCV多聚蛋白的表达调控机制提供理论信息。

1 材料与方法

1.1 选取的基因样本 文中所用18条EMCV完整编码基因的相关信息来源于GenBank(<http://www.ncbi.nlm.nih.gov>),序列号分别为:AF356822、DQ464062、DQ464063、FJ604852、FJ604853、FJ897755、HM641897、DQ517424、EU78-0148、EU780149、DQ288856、X87335、AY296731、M37588、M22457、X74312。这些序列将在整体上体现EMCV病毒基因组组成上的遗传特性,从而提炼出EMCV的生物多样性。

1.2 基因密码子相对使用概率 与密码子使用概率计算相关的统计量有密码子使用频次、绝对密码子使用概率、相对密码子使用概率等。其中,密码子使用相对概率(relative synonymous codon usage, RSCU)是衡量密码子使用偏性的方法之一。RSCU是对于某一特定的密码子在编码对应氨基酸的同义密码子间的相对概率。它排除了氨基酸组成对密码子使用的影响,在密码子的使用无偏嗜性的情况下,该密码子的RSCU值接近1.0。若某一密码子的RSCU值大于1.0,表示此密码子是使用相对较多的密码子;当RSCU值小于1.0时,表示此密码子是使用相对较少的密码子。RSCU的优点是计算简便,且直接反映了密码子使用的偏嗜性^[5]。文中所涉及密码子使用偏性的分析都以RSCU为衡量标准。

基金项目 甘肃省农业支撑计划(No. 1104NKCA082)。

作者简介 张瑜(1983-),女,陕西米脂人,助理工程师,硕士研究生,从事动物医学方面的研究。*通讯作者,教授,从事兽医公共卫生学方面的研究, E-mail: sunxl@gsau.edu.cn。

收稿日期 2013-07-25

1.3 翻译起始区与终止区的密码子使用偏嗜性 为研究 EMCV 编码区的翻译起始和终止区同义密码子使用的特点,文中借鉴一种测评局部序列中同义密码子使用偏嗜性的方法^[6],分别利用所建立的计算公式计算此种病毒前 240 个密码子和末尾 240 个密码子区域所组成的翻译起始和终止区内部各同义密码子的使用偏嗜性。

$$Y = \log_n \left(\frac{f_n/F_n}{f/F} \right)$$

在此公式中, f_n 代表翻译起始或终止区域中特定的同义密码子出现的次数; F_n 代表特定同义密码子家族中所有成员在翻译起始或终止区域中出现的次数。 f 代表特定同义密码子在整体编码基因中出现的次数; F 代表特定密码子家族中所有成员在整体编码基因中出现的次数。 Y 值是衡量特定同义密码子在规定区域内使用偏嗜性的参考指标^[7],当 Y 值大于 0.2 时,对应同义密码子在目标区域内的使用偏嗜性为正偏向;当 Y 值小于 -0.2 时,对应同义密码子在目标区域内的使用偏嗜性为负偏嗜。

2 结果与分析

2.1 同义密码子在 EMCV 中使用的模式 研究发现 EMCV 59 种同义密码子的使用模式并不是呈现出一种随机性和均一性,而是呈现出一种特定的模式(表 1)。在所有 RSCU 值大于 1.0 的密码子中,密码子第 3 位以 A/U 结尾的共有 16 个,而以 G/C 结尾的共有 11 个;而对于所有 RSCU 值小于 1.0 的密码子,密码子第 3 位以 A/U 结尾的共有 14 个,而以 G/C 结尾的共有 16 个。总体上来看,EMCV 对于以 G/C 为第 3 位结尾的密码子使用的倾向性略小于以 A/U 结尾的密码子。

2.2 EMCV 翻译起始区域的密码子使用偏嗜性 由图 1 可看出,大多数同义密码子在 EMCV 翻译起始序列区内的使用是有明显的正或负偏嗜性的,同时,也存在一部分 RSCU 值处于 1.0 附近的同义密码子在相应区域内的使用不具有明显的正或负偏嗜性。这可能暗示了:同义密码子在影响 EMCV 相关基因的翻译起始效率的过程中扮演着不同的角色。按照 RSCU 值小于 0.6 且 Y 值大于 0.2 的标准来看,虽然 UUA (Leu)、CGA (Arg)、UCG (Ser)、ACG (Thr) 和 GCG (Ala) 在 EMCV 整体编码区内的使用偏嗜性为负偏嗜,但这些密码子在病毒的翻译起始序列区内呈现的模式为正偏嗜。按照 RSCU 值大于 1.0 且 Y 值小于 0.2 的标准来比较分析,虽然 UUG (Leu)、CUG (Leu)、AUU (Ile)、GUG (Val)、CCU (Pro)、CCC (Pro)、ACU (Thr)、ACC (Thr)、GCC (Ala)、UAU (Tyr)、CAU (His)、CAG (Gln)、AAG (Lys)、GAG (Glu)、AGA (Arg)、AGG (Arg) 和 GGA (Gly) 在 EMCV 整体编码区内的出现频率呈正偏嗜,但这些密码子在病毒的翻译起始序列区内的使用频率呈明显的负偏嗜。总体来看,EMCV 翻译起始序列区内倾向于避开使用那些在整体翻译编码区内出现频率高的密码子,而那些在整体翻译起始编码区内出现频率很低的密码子却偏向于出现在翻译起始区内,这暗示了稀有密码子在 EMCV 翻译起始的调控过程中发挥着重要作用。

表 1 EMCV 的同义密码子使用模式

密码子(氨基酸)	RSCU 值	密码子(氨基酸)	RSCU 值
UUU(F)	1.11	GCU(A)	1.21
UUC(F)	0.89	GCC(A)	1.44
UUA(L)	0.45	GCA(A)	0.84
UUG(L)	1.43	GCG(A)	0.52
CUU(L)	0.96	CAA(Q)	0.83
CUC(L)	0.99	CAG(Q)	1.17
CUA(L)	0.77	AAU(N)	1.18
CUG(L)	1.41	AAC(N)	0.82
AUU(I)	1.56	AAA(K)	0.85
AUC(I)	0.70	AAG(K)	1.15
AUA(I)	0.73	GAU(D)	1.18
GUU(V)	0.92	GAC(D)	0.82
GUC(V)	0.95	GAA(E)	0.93
GUA(V)	0.64	GAG(E)	1.07
GUG(V)	1.49	UGU(C)	0.99
UCU(S)	1.54	UGC(C)	1.01
UCC(S)	1.32	CGU(R)	0.83
UCA(S)	1.43	CGC(R)	0.69
UCG(S)	0.54	CGA(R)	0.34
AGU(S)	0.59	CGG(R)	0.62
AGC(S)	0.58	AGA(R)	2.52
CCU(P)	1.06	AGG(R)	1.00
CCC(P)	1.11	GGU(G)	1.06
CCA(P)	1.33	GGC(G)	0.91
CCG(P)	0.5	GGA(G)	1.26
UAU(Y)	1.01	GGG(G)	0.77
UAC(Y)	0.99	ACU(T)	1.26
CAU(H)	1.22	ACC(T)	1.28
CAC(H)	0.78	ACA(T)	1.14
		ACG(T)	0.32

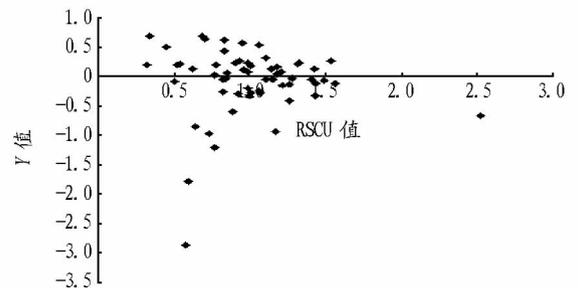


图 1 EMCV 翻译起始序列区的密码子使用模式

2.3 EMCV 的翻译终止区域的密码子使用偏嗜性 图 2 显示了密码子在 EMCV 翻译终止序列区内的使用偏嗜性。有趣的是,研究同样观察到大多数同义密码子在 EMCV 翻译终止序列区内的使用是有明显正或负偏嗜性的,这也同样表明了此区域同义密码子使用具有非随机性和非均一性的遗传特点。按照 RSCU 值小于 0.6 且 Y 值大于 0.2 的标准来看,虽然 UUA (Leu)、UCG (Ser)、GCG (Ala) 和 CGA (Arg) 在 EMCV 整体编码区内的使用偏嗜性为负偏嗜,但这些密码子在病毒的翻译终止序列区内出现的模式呈正偏嗜。按照 RSCU 值大于 1.0 且 Y 值小于 0.2 的标准来看,虽然 CUG (Leu)、AUU (Ile)、UCU (Ser)、UCC (Ser)、CCC (Pro)、ACC (Thr)、ACA (Thr)、GCU (Ala)、GCC (Ala)、CAG (Gln)、UGC (Cys)、AGA (Arg) 和 GGA (Gly) 在 EMCV 整体编码区内的出

现频率呈正偏嗜,但这些密码子在病毒的翻译终止序列区内的使用频率呈负偏嗜。总体来看,EMCV 翻译终止序列区内同样倾向于避开使用那些在整体翻译编码区内出现频率高的密码子,而那些在整体翻译起始编码区内出现频率很低的密码子却偏向于出现在翻译起始区内,这暗示了稀有密码子在 EMCV 翻译终止的调控过程中也发挥着重要作用。

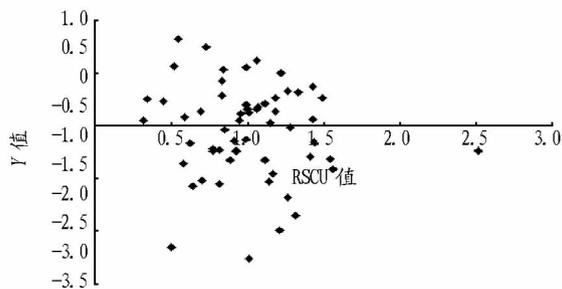


图2 EMCV 翻译终止序列区的密码子使用模式

3 讨论

在自然界中,氨基酸序列是由其对应的 mRNA 序列指导核糖体进行翻译合成的,但由于遗传密码的简并性,mRNA 的编码序列与相应的氨基酸并非简单地一一对应关系,而是多种密码子可产生相同的氨基酸物质。在此种物质基础上,核苷酸编码序列所包含的遗传信息就要远多于其对应产生的氨基酸序列所携带的遗传信息^[8]。氨基酸多肽链在核糖体内部的合成过程中,由于同义密码子与其互补的反密码子的特性及所处细胞环境的不同将显著影响氨基酸多肽链的合成状态。对大肠杆菌编码区不同位置上同义密码子使用情况的研究表明,多数氨基酸的密码子用语在翻译起始区变化显著,只有少数氨基酸在翻译终止区有微弱变化^[9]。在研究一些病毒同义密码子使用模式的过程中发现,同义密码子用语与氨基酸序列上下文间存在密切的相关性,尤其是翻译起始序列区域的密码子用语影响最为显著^[10-13]。

对不同 EMCV 的翻译起始序列区域和翻译终止序列区域的密码子使用偏嗜性的分析对于探讨病毒在宿主细胞中的翻译调控机制是十分必要的。目前,在对真核生物相关基因的研究中发现,其开放阅读框起始端的密码子使用偏嗜性将显著影响对应基因在相应宿主细胞中的表达效率,从而导致基因所翻译的蛋白质可准确地从无生物学活性的新生肽段转变为具有正常生物学功能的蛋白质。这一现象是由与出现在翻译起始序列区域的同义密码子所对应 tRNA 的丰度所决定的。因为同义密码子所对应的 tRNA 在细胞中的含量是呈正相关的,所以当一些对应 tRNA 低丰度的同义密码子出现在翻译起始序列区的频次较高时,将显著影响核糖体向下游编码序列的移动速率,从而有效调控相应基因的翻译效率。此外,在关注翻译起始序列区同义密码子使用偏嗜性的同时,该研究还分析了 EMCV 翻译终止区域同义密码子使用的偏嗜性,并发现一些使用偏嗜性较低的同义密码子也倾向于出现在此处。这也就说明,“稀有密码子调控”假说同样可应用于翻译终止序列区,即在新生肽段将要翻译表达结束的过程中,稀有密码子同样会导致核糖体扫描速率的降低,从

而为新生肽段正确折叠成为具有生物学活性的蛋白质创造条件^[14-16]。

总之,对于 EMCV 来说,在翻译起始和终止区域的同义密码子使用偏嗜性调控其多聚蛋白质的表达和加工成熟是具有重要意义的。那些大量存在于这 2 种区域的稀有密码子可利用“稀有密码子调控”假说来解释。这些稀有的同义密码子将对多聚蛋白质的表达产生负调控效应。由此可见,“稀有密码子调控”假说不仅适用于细菌等低等微生物,而且对 RNA 病毒相关编码序列在核酸水平上的翻译调控同样具有十分重要的意义。更深入的一层是,这些有趣的现象可以为研究 EMCV 的进化特点和表达调控提供一条新颖的研究思路。

参考文献

- [1] 盖新娜,杨汉春,郭鑫,等. 猪脑心肌炎病毒的分离与鉴定[J]. 畜牧兽医学报,2007,38(1):59-65.
- [2] ZHANG G Q, GE X N, GUO X, et al. Genomic analysis of porcine encephalomyocarditis virus strains isolated in China[J]. Arch Virol, 2007, 152(6):1209-1213.
- [3] 赵婷,张家龙,盖新娜,等. 猪源和鼠元脑心肌炎病毒分离株基因组的比较分析[J]. 畜牧兽医学报,2009,40(6):873-878.
- [4] ZHOU J H, ZHANG J, DING Y Z, et al. Characteristics of codon usage bias in two regions downstream of the initiation codon of foot-and-mouth disease virus[J]. Bio Systems, 2010, 101(1):20-28.
- [5] ZHOU J H, ZHANG J, CHEN H T, et al. Analysis of synonymous codon usage in foot-and-mouth disease virus[J]. Vet Res Commun, 2010, 34:393-404.
- [6] ZHOU J H, ZHANG J, CHEN H T, et al. The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus[J]. Infection, Genetics and Evolution, 2011, 11:1815-1819.
- [7] ZHOU T, GU W, MA J, et al. Site discrepancy of synonymous codon usage in SARS coronavirus and other viruses in Coronaviridae[J]. Journal of Southeast University, 2005, 21(2):203-206.
- [8] 李炜疆,宋江宁. 核酸序列是否包含更多蛋白质结构信息[J]. 无锡轻工业大学学报,2001,21(6):652-655.
- [9] 李炜疆. 同义密码子用语的位置依赖[J]. 生物物理学报,2001(17):529-534.
- [10] KARLIN S, MRAZEK J. What drives codon choices in human genes[J]. J Mol Biol, 1996, 262:459-472.
- [11] GAVRILIN G V, CHERKASOVA E A, LIPSKEYA G Y. Evolution of circulating wild poliovirus and of vaccine-derived poliovirus in an immunodeficient patient: a unifying model[J]. J Virol, 2000, 74(16):7381-7390.
- [12] BURNS C C, SHAW J, CAMPAGNOLI R, et al. Modulation of poliovirus replicative fitness in Hela cells by deoptimization of synonymous codon usage in the capsid region[J]. J Virol, 2006, 80(7):3259-3272.
- [13] LI B, XIAO S, WANG Y, et al. Immunogenicity of the highly pathogenic porcine reproductive and respiratory syndrome virus GP5 protein encoded by a synthetic ORF5 gene[J]. Vaccine, 2009, 27:1957-1963.
- [14] CHEN G F, INOUE M. Suppression of the negative effect of minor arginine codons on gene expression: preferential usage of minor codons with the first 25 codons of the Escherichia coli genes[J]. Nucleic Acids Res, 1990, 18:1465-1473.
- [15] CHEN G F, INOUE M. Role of the AGA/AGG codons, the rarest codons in global gene expression in Escherichia[J]. Genes Dev, 1994, 8:2641-2652.
- [16] OHNO H, SAKAI H, WASHIO T, et al. Preferential usage of some minor codons in bacteria[J]. Gene, 2001, 276:107-115.
- [17] 房拴柱,杨贵清,刘培德. 脑心肌炎的防控[J]. 畜牧与饲料科学, 2010, 31(11-12):161.
- [18] 马晓霞,冯玉萍,赵永清,等. 口蹄疫病毒密码子使用偏性的位点差异(英文)[J]. 华北农学报,2013(2):32-37.
- [19] 梁菲菲. 密码子偏性的影响因素及研究意义[J]. 畜牧与饲料科学, 2010, 31(1):118-119.