Journal of Biotechnology 181 (2014) 31–34

Contents lists available at ScienceDirect

Journal of Biotechnology

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

Short communication

Comparative analysis of internal ribosomal entry sites as molecular tools for bicistronic expression

Eldem Sadikoglou, Elena Daoutsali, Eleni Petridou, Maria Grigoriou, George Skavdis*

Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, Building 10, Dragana, Alexandroupolis 68100, Greece

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 2 March 2014
Accepted 25 March 2014
Available online 4 April 2014

Keywords:
Bicistronic expression
IRES
EMCV
VCIP
Gtx9

ABSTRACT

Internal ribosomal entry sites (IRESes) are sequences that drive cap-independent translation. They are found in some viral and cellular transcripts and they have been extensively used in both basic and applied research for the translation of two or more polypeptides from a single mRNA molecule in eukaryotic cells. Although the most widely used IRES comes from the encephalomyocarditis virus (EMCV), several other viral and cellular IRES elements have been identified and successfully used, including those of the human *VCIP* gene and the mouse *Gtx* gene. In this report we have compared the EMCV IRES with the VCIP and the Gtx IRESes, and we provide evidence that by using the EMCV IRES much higher levels of second cistron expression can be achieved.

Since the initial characterization of the encephalomyocarditis virus (EMCV) internal ribosomal entry site (IRES) (Palmenberg et al., 1984; Parks et al., 1986), this element has been successfully used in numerous eukaryotic expression vectors for the coexpression of two polypeptides from a single mRNA molecule. However, over the last years several reports have appeared in the literature showing that in some bicistronic mRNAs, the expression levels of the second cistron are unexpectedly low (Hennecke et al., 2001; Mizuguchi et al., 2000). In an attempt to interpret these results, Yury Bochkov and Ann Palmenberg suggested that the compromised expression of the second cistron, at least in some cases, could be attributed to the use of attenuated versions of the EMCV IRES (Bochkov and Palmenberg, 2006). The EMCV IRES contains 12 AUG triplets. It has been shown that the 11th AUG (AUG^{11}) is the one primarily used for the initiation of protein synthesis (Davies and Kaufman, 1992; Smith, 1973). A modified EMCV IRES widely found in bicistronic vectors, in which the authentic AUG^{11} has been replaced by a HindIII recognition sequence (thus forcing the translation to start from the downstream located AUG^{12}), drives the expression of the second cistron at levels 10-fold lower than those of the wild type element (Martin et al., 2006). A different modification based on the insertion of an ACC triplet before the AUG^{11} has also been shown to negatively affect the EMCV IRES (Qiao et al., 2002). Another factor that influences the efficiency of the EMCV IRES is the presence of an extra A nucleotide in a short oligo(A) region located 60 nucleotides upstream of the AUG^{11}. Although this deviation from the wild type sequence results in an approximately twofold reduction of the second cistron expression levels, some of the vectors that harbor the EMCV IRES have 7 instead of 6 A residues in this oligo(A) region (Bochkov and Palmenberg, 2006).

Two cellular IRESes that have been reported to outperform in a bicistronic context the widely used EMCV IRES are those of the mouse *Gtx* gene and the human *VCIP* gene that code for a homeobox transcription factor and an adhesion protein respectively (Chappell et al., 2000; Licursi et al., 2011; Wang et al., 2005). However, a closer examination of the relevant literature reveals that it is not clear exactly which version of the EMCV IRES was compared with the Gtx IRES. Moreover, the data published by Licursi et al. indicate that the EMCV IRES is one of the less efficient elements among the 13 IRESes tested, driving second cistron expression at levels often lower than those detected with an empty-vector control construct. Since these results are rather surprising for an element that has been successfully used by many laboratories and for many years, we carefully examined the EMCV IRES sequences used by Licursi et al. based on the information provided by these authors. As it is shown in Fig. 1, it appears that the EMCV IRES they used, should be a severely attenuated version of the wild type sequence (from now on called $EMCV^{att}$ IRES) because it is missing 4 out of 9 nucleotides between the authentic AUG^{11} and the AUG^{12}. Since it is well documented that the AUG^{11} is primarily used for the initiation of protein

* Corresponding author. Tel.: +30 2551030626; fax: +30 2551030657.
E-mail address: gskavdis@mbg.duth.gr (G. Skavdis).

http://dx.doi.org/10.1016/j.jbiotec.2014.03.033
0168-1656/

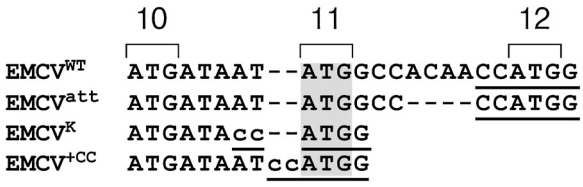

Fig. 1. Alignment of the four EMCV IRES variants used in this study. For each element the DNA sequence surrounding the authentic ATG[11] initiation codon (shaded in gray) is presented. The positions of ATG[10] and ATG[12] are also indicated. The NcoI site used for the cloning of the second cistron (firefly *luc* gene) is underlined. Nucleotides depicted in lowercase indicate substitutions (EMCV^{K}) or insertions (EMCV^{+cc}) relatively to the wild type sequence (EMCV^{wt}).

synthesis (Davies and Kaufman, 1992; Smith, 1973), apparently most of the second cistron protein produced by the EMCV^{att} IRES is out of frame.

To clarify these issues we prepared a series of constructs carrying the wild type EMCV IRES (EMCV^{wt}), the EMCV^{att}, the VCIP and the Gtx IRESes. As a first cistron we used the *tdTomato* gene (Shaner et al., 2004) that codes for a fluorescent protein while as a second cistron we used the gene of the firefly luciferase (*luc*). In order to be able to directly compare the expression levels achieved from the bicistronic mRNAs to those of the corresponding monocistronic mRNAs, we also generated two constructs that express either the *tdTomato* or the *luc* gene alone. However, when the intact EMCV^{wt} IRES is used, the second open reading frame is cloned into a naturally occurring NcoI site surrounding the AUG[12]. Since the translation initiates mainly at the AUG[11], the protein encoded by the second cistron carries an amino-terminal tail consisting of the amino acid residues Met-Ala-Thr-Thr. This tail could affect the activity of the luciferase synthesized under the EMCV^{wt} IRES, leading to erroneous interpretation of the results. To address this possibility we also prepared a monocistronic construct that expresses the tailed version of the luciferase protein (t-luc).

The aforementioned constructs were used to transfect a human and a mouse cell line, namely HEK-293 and Neuro-2A. As it is shown in Fig. 2, the standard luciferase and the t-luc are expressed at very similar levels from the corresponding monocistronic constructs indicating that the results obtained from the EMCV^{wt} IRES can be directly compared with these obtained from the other IRESes. In both cell lines, the EMCV^{wt} IRES drives strong luciferase expression, at levels approximately 10–30 fold higher than the EMCV^{att} IRES. The luciferase expression levels driven by the constructs carrying the VCIP and the Gtx IRESes were as low or even lower than those of EMCV^{att} IRES construct. We also noticed that in the bicistronic constructs the expression of the first cistron is lower than that of the corresponding monocistronic construct, an observation that has been previously described (Bochkov and Palmenberg, 2006; Qiao et al., 2002). In principle, this phenomenon can be attributed to differences in the mRNA levels, and/or to reduced cap-dependent translation efficiency in the bicistronic constructs. To explore these possibilities we isolated total RNA from transfected HEK-293 cell, converted it to cDNA and used it as template in real time PCR with primers specific for the 3′ untranslated region that is common to all the mRNAs studied. As it is shown in Fig. 3 the mRNA levels of the bicistronic constructs are lower than those of the monocistronic constructs. The mRNA of the tdT construct accumulates at levels approximately three times higher than the mRNA of the EMCV^{wt} IRES construct. This observation suggests that the approximately four times difference at the tdT protein levels between the two constructs (Fig. 2) can be largely attributed to differences in the mRNA levels and to a lesser extend to reduced cap-dependent translation efficiency through, for example, a competition between the cap-dependent and the IRES-dependent translation machineries (Svitkin et al., 2005). All bicistronic constructs had very similar mRNA levels, indicating that the differences between them at the

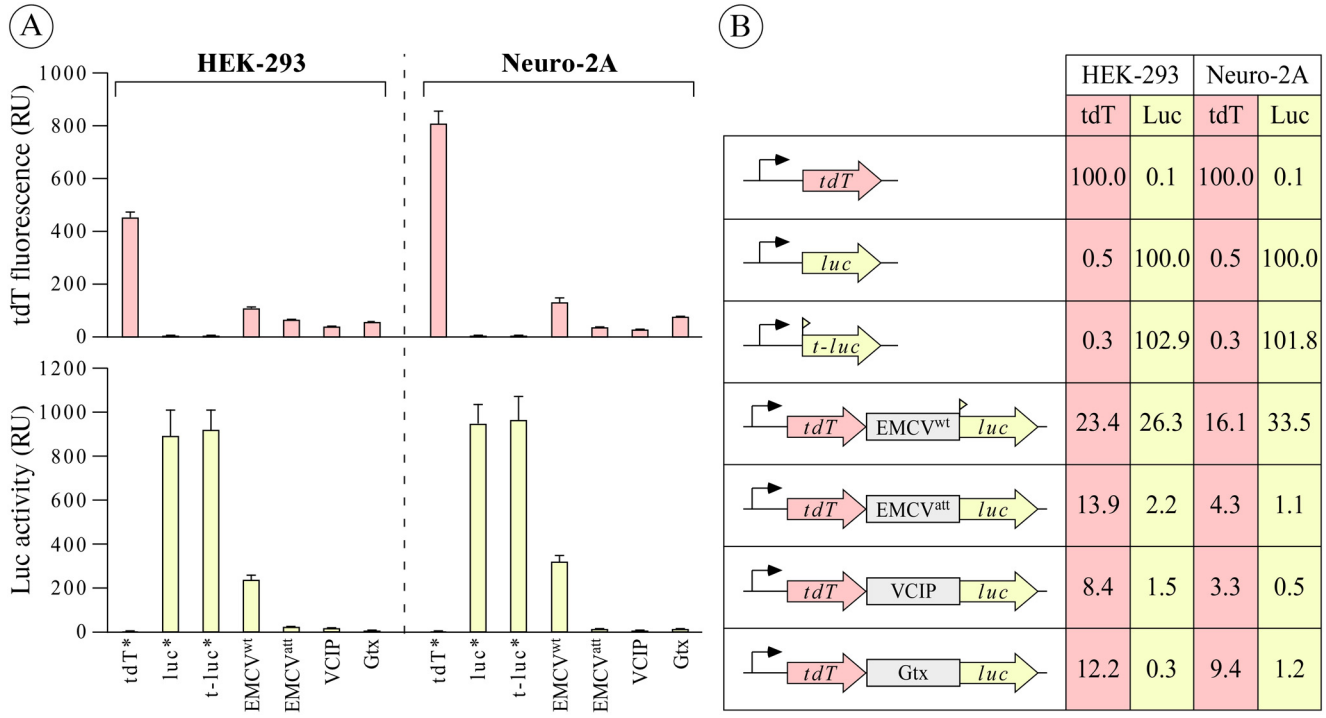

Fig. 2. The EMCV^{wt} IRES can mediate higher levels of expression than the Gtx9 and the VCIP IRESes. (A) Histograms showing tdTomato fluorescence levels and luciferase activity measured in HEK-293 and Neuro-2A after transfection with the indicated monocistronic (marked with a star) or bicistronic constructs. Error bars represent standard error. RU: relative units. (B) Table showing a graphical representation of the constructs used in this experiment and the expression levels of the reporter genes as the percentage over the expression levels of the corresponding monocistronic constructs. A flag indicates the *t-luc* gene. The bent arrow indicates the human CMV promoter. For more experimental details please see Supplementary data.