Conservation of *cis*-Regulatory Element Controlling Timely Translation in the 3'-UTR of Selected Mammalian Maternal Transcripts

Hyun-Joo Lee[2], Yoonki Lim[2], Sang-Ho Chang[2], Kwansik Min[1], Ching-Tack Han[3] and Sue-Yun Hwang[1]*

[1]Graduate School of Biotechnology, Environmental and Information Technology, [2]Department of Bioengineering, Hankyong National University, Ansung 456-749, Korea, [3]Department of Life Science, Sogang University, Seoul 121-742, Korea

Abstract

The earliest stages of mammalian embryogenesis are governed by the activity of maternally inherited transcripts and proteins. Cytoplasmic polyadenylation of selected maternal mRNA has been reported to be a major control mechanism of delayed translation during preimplantation embryogenesis in mice. The presence of *cis*-elements required for cytoplasmic polyadenylation (e.g., CPE) can serve as a useful tag in the screening of maternal genes partaking in key functions in the transcriptionally dormant egg and early embryo. However, due to its relative simplicity, UA-rich sequences satisfying the canonical rule of known CPE consensus sequences are often found in the 3'-UTR of maternal transcripts that do not actually undergo cytoplasmic polyadenylation. In this study, we developed a method to confirm the validity of candidate CPE sequences in a given gene by a multiplex comparison of 3'-UTR sequences between mammalian homologs. We found that genes undergoing cytoplasmic polyadenylation tend to create a conserved block around the CPE, while CPE-like sequences in the 3'-UTR of genes lacking cytoplasmic polyadenylation do not exhibit such conservation between species. Through this cross-species comparison, we also identified an alternative CPE in the 3'-UTR of tissue-type plasminogen activator (tPA), which is more likely to serve as a functional element. We suggest that verification of CPEs based on sequence conservation can provide a convenient tool for mass screening of factors governing the earliest processes of mammalian embryogenesis.

*Corresponding author: E-mail dutuya@hknu.ac.kr
Tel +82-31-670-5424, Fax +82-31-670-5419

Introduction

In all metazoan species, maternal mRNAs are synthesized and stored in the egg cytoplasm in a translationally dormant form. During oocyte maturation and early development, timely activation of these mRNAs promotes synthesis of key proteins, orchestrating the second meiotic division, fertilization, and cleavage (Davidson, 1986). A mechanism unique to this period is known to control the timely translation of selected maternal messages by modulating the length of the poly(A) tail in the cytoplasm. (Richter, 1999; Mendez and Richter, 2001). This secondary elongation of the poly(A) tail is termed cytoplasmic polyadenylation, to distinguish it from the original process occurring in the nucleus. The essential nature of this process has been well demonstrated in experiments in which impairment of cytoplasmic polyadenylation blocked timely translation of corresponding genes and normal embryonic development (Gebauer *et al*., 1994; Lieberfarb *et al*., 1996).

The *cis*-elements necessary for cytoplasmic polyadenylation (e.g., CPE) are located in the vicinity of the nuclear polyadenylation signal, and they trigger elongation of the poly(A) tail in the egg and zygote (Stebbins-Boaz *et al*., 1997; de Moore and Richter, 2001). Since the earliest period of embryonic development is still governed by maternally inherited factors, presence of a CPE in the 3'-UTR serves as a useful marker for genes partaking in key functions at the onset of mammalian embryogenesis (Ohsugi *et al*., 1996; Oh *et al*., 1997; Read and Norbury, 2002). However, due to its relative simplicity, the UA-rich sequences satisfying the canonical rule of the mammalian CPE (usually UUUUAU) are often found in the 3'-UTR of genes that do not undergo cytoplasmic polyadenylation (Hwang *et al*., 2001; Lee *et al*., 2006). Unfortunately, experimental verification of poly(A) elongation is not feasible for mammalian maternal mRNAs due to the inaccessibility of oocytes and early embryos.

While the nucleotide sequence of a typical mammalian gene shows strict conservation between homologs within

the ORF, the degree of similarity tends to decrease sharply in the 3'-UTR. Yet there appear to be limited regions of conserved blocks around known 3'-UTR elements, indicating their functional importance. Examples of such cases include the hexanucleotide nuclear polyadenylation signal (AAUAAA or AUUAAA) and the instability elements of cytokines (Seko *et al.*, 2006; Paschoud *et al.*, 2006). In this study, we examined whether the 3'-UTR sequences around the CPE also create a conserved island across multiple mammalian species. We found that sequence conservation around a potential CPE reflects the validity of this *cis*-element, which may provide a convenient measure to distinguish genes involved in the onset of mammalian embryogenesis.

Materials and methods

Nucleotide sequences in the 3' end of analyzed genes were obtained through GenBank. The GenBank identification numbers of examined sequences are listed in Table 1. When sequence information for the extreme 3' end of a

Table 1. GenBank identification number of analyzed sequences.

Gene	Species	ID number *	Area used in alignment (shown by nucleotide number)
E-cadherin	*H. sapiens*	Z13009	4609~4778
	M. musculus	X06115	4199~4356
	S. scrofa	DV898711*	94~266
	B. taurus	DT861152 *	522~685
α-catenin	*H. sapiens*	NM_001903	3349~3765
	M. musculus	NM_009818	3299~3680
	S. scrofa	AY550072	569~929
	B. taurus	NM_001035366	3313~3693
β-catenin	*H. sapiens*	NM001904	3490~3669
	M. musculus	NM_007614	3379~3552
	S. scrofa	NM_214367	2996~3174
	B. taurus	NM_001076141	3145~3322
γ-catenin	*H. sapiens*	NM_002230	3348~3490
	M. musculus	BC040757	3016~3193
	S. scrofa	NM_214323	3066~3224
	B. taurus	NM_001004024	2941~3095
β-actin	*H. sapiens*	BC002409	1597~1815
	M. musculus	NM_007393	1684~1982
	S. scrofa	AY550069	1643~1862
	B. taurus	AY141970	1567~1796
tPA	*H. sapiens*	M18182	2297~2461
	M. musculus	BC011256	2358~2530
	S. scrofa	AK234864*	2312~2488
	B. taurus	NM_174146*	1567~1796
Melk	*H. sapiens*	NM_014791	2270~2442
	M. musculus	X95351	2244~2471
	S. scrofa	AJ655298*	251~414
	B. taurus	XM_598320*	2171~2343

*EST clones representing the 3' end of the gene.

gene was not available from annotated entries, a contig of the missing regions was generated using overlapping EST fragments.

Interspecies alignment of 3'-UTR sequences was performed with homologs from the genomes of human (*Homo sapiens*), mouse (*Mus musculus*), pig (*Sus scrofa*), and cow (*Bos taurus*) (Choi *et al.*, 2006). Quadruple alignments of sequences were performed by the ClustalW program of the MEGA3 software using standard parameters (Kumar *et al.*, 2004).

Results and Discussion

Conservation of 3'-UTR sequences around CPEs of mammalian cell adhesion molecules

Like other mouse proteins controlling the earliest developmental process, cell adhesion molecules involved in the compaction of the 8-cell stage embryo are produced by timely translation of stored maternal messages (Ohsugi *et al.*, 1996). Compacting embryos showed strong expression of E-cadherin, α-catenin, and β-catenin at the junction of blastomeres. It has been demonstrated that the poly(A) tail of these transcripts is elongated by several hundred nucleotides before the activation of protein synthesis (Oh *et al.*, 2000). Potential CPE sequences meeting the canonical consensus from mouse genes known to undergo cytoplasmic polyadenylation are present in the 3'-UTRs of E-cadherin, α-catenin, and β-catenin (Gebauer *et al.*, 1994; Richards *et al.*, 1993). γ-catenin (also known as plakoglobin) constitutes the third component of the cadherin-catenin complex in certain cell types, but γ-catenin protein is not detected in the compaction stage of the embryo (Ohsugi *et al.*, 1996). Interestingly, the 3'-UTR of mouse γ-catenin does not have any UA-rich sequences satisfying the CPE consensus sequence.

In this study, we compared the distal 3' end of transcripts encoding the cadherin-catenin complex by a multiplex alignment of four mammalian homologs. From the known cases of mouse and *Xenopus* CPEs, we built a broader consensus sequence defined as any UA-rich stretch resembling (A)UUUU(U)AU and residing within approximately 100 nucleotides from the nuclear polyadenylation signal. As shown in Fig. 1, the 3' end of E-cadherin, α-catenin, and β-catenin all possess hexanucleotide nuclear polyadenylation signals (either AAUAAA or AUUAAA), perfectly conserved among the four mammalian genes examined (boxes with broken lines). Equally well conserved in mammals were sequences around the candidate CPE sequences of these genes (box with solid lines). We also noticed additional UA-rich sequences in the 3'-UTR of α-catenin that were strictly conserved in all four species (Fig. 1 B, underlined),

Fig. 1. Comparisons of the 3' end sequences between mammalian homologs of cell adhesion molecules. **A:** E-cadherin, **B:** α-catenin, **C:** β-catenin. Boxes with broken lines indicate the hexanucleotide nuclear polyadenylation signal. Previously identified CPE sequences are marked by boxes with solid lines. Additional CPE-like sequences in the 3' end of α-catenin and β-catenin are marked with double-underlines.

Fig. 2. Comparisons of the 3'-end sequences between mammalian homologs of maternal transcripts that avoid cytoplasmic polyadenylation. **A:** β-actin, **B:** *Melk*, **C:** γ-catenin (plakoglobin). Boxes with broken lines indicate the hexanucleotide nuclear polyadenylation signal. Candidate CPE sequences are marked by boxes with solid lines.

although it showed slight deviation from the known consensus CPE. This result implies the functional importance of CPEs in the timely translation of corresponding proteins in the early embryos.

For β-catenin, the validity of the downstream UUUUAAUU (Fig. 1 C, underlined) had been in question because it is located past the nuclear polyadenylation signal. However, cyclin B1 mRNA of *Xenopus* has a functional CPE that overlaps the nuclear polyadenylation signal (de Moore and Richter, 1999). In the present study, conservation of the UUUUAAUU sequence of β-catenin adds to the possibility that CPEs may reside downstream of the nuclear polyadenylation signal as well.

Absence of cross-species conservation around CPE-like sequences in maternal transcripts that do not undergo cytoplasmic polyadenylation

Timely polyadenylation of stored maternal messages in the cytoplasm can be monitored by Northern blot analysis of oocytes and embryos (Oh *et al*., 2000). In this analysis,

RNaseH digestion after oligo-dT hybridization abolished the upward shift of the Northern band, indicating that the size increase was contributed by the addition of the poly(A) tail. According to this method, transcripts of mouse β-actin and *Melk* do not undergo cytoplasmic polyadenylation despite their abundance in the egg and early embryo (Oh *et al*., 2000). However, the 3'-UTR of these messages contain several UA-rich sequences observing the canonical rule of the CPE (Fig. 2). One explanation for this dilemma might be the fact that CPE-like sequences of β -actin and *Melk* transcripts are located beyond the cut-off distance from the nuclear polyadenylation signal, which is generally believed to be approximately 100 nucleotides. Interestingly, these CPE-like sequences are not conserved between mammalian homologs (Fig. 2 A and B; see boxes with solid lines). We also noticed a similar deviation in the

3'-UTR alignment of γ-catenin, which is not actively translated in the mouse egg and embryo. There are no CPE-like sequences in the 3' end of mouse γ-catenin, but the other three mammals contain UUUUAU motifs (underlined) about 60 nucleotides away from the nuclear polyadenylation signal (Fig. 2 C). Such divergence may reflect the fact that the corresponding area had not been under the force of evolutionary conservation, presumably because they do not participate in the cytoplasmic polyadenylation. Taken together, our results suggest that examination of interspecies conservation is one way to determine the authenticity of candidate CPEs.

Suggestion for an alternative CPE in tPA via cross-species alignment of 3'-UTR

The protease tissue-type plasminogen activator (tPA) was one of the first vertebrate genes shown to undergo cytoplasmic polyadenylation in developing oocytes (Huarte *et al*., 1987). Since the presence of tPA protein can be determined by zymography, it is possible to monitor its translation from a small number of mouse eggs. The 3'-UTR of mouse tPA contains a variety of CPE-like sequences, including two AUUUUAAU stretches. A series of antisense inhibition tests demonstrated that an oligonucleotide targeting the downstream AUUUUAAU abolished the timely translation of tPA after ovulation (Richards *et al*., 1993), suggesting that this element is involved in the translational activation of stored maternal messages in the egg. To our surprise, both upstream and downstream AUUUUAAU sequences were not conserved between mammals (Fig. 3, boxes with solid lines). Instead, a strict conservation was noticed from another UA-rich sequence, UAUUUAUA (box with solid lines), juxtaposed to the downstream AUUUUAAU sequence. Interestingly, the antisense oligo inhibiting tPA translation in the experiments by Richards *et al*. included the UAUUUAUA sequence as well. According to what has been observed from maternal transcripts examined in this study, the conserved UAUUUAUA sequence, instead of

AUUUUAAU, most likely serves as a functional CPE in oocytes. It would be intriguing to test whether disrupting UAUUUAUA alone can inhibit the translation of tPA in the egg.

Taken together, sequence conservation between related species may designate the validity of given CPE-like elements. Such a measure could be particularly useful in mammals wherein direct verification of cytoplasmic polyadenylation is not practical due to the paucity of egg and early stage embryos. We suggest that screening of CPE-containing genes using this method will contribute to the understanding of cellular events occurring at the onset of mammalian embryogenesis.

Acknowledgments

References

Choi, H.S., Kim, D.W., and Ryu, W. (2006). A pattern summary system using BLAST for sequence analysis. *Genomics & Informatics* 4. 173-181.

Davidson, E.H. (1986). Gene activity in early development. Academic Press. New York.

de Moore, C. and Richter, J. (1999). Cytoplasmic polyadenylation elements mediate kasking and unmasking of cyclin B1 mRNA. *EMBO J.* 18. 2294-2303.

de Moore, C. and Richter, J. (2001). Translational control in vertebrate development. *Int Rev Cytol.* 203. 567-608.

Gebauer, F., Xu, W., Cooper, J., and Richter, J. (1994). Translational control by cytoplasmic polyadenylation of *c-mos* mRNA is necessary for oocyte maturation in the mouse. *EMBO J.* 13. 5712-5720.

Gebauer, F., and Richter, J. (1996). Mouse cytoplasmic polyadenylation element binding protein: An evolutionary conserved protein that interact with the cytoplasmic polyadenylation elements of *c-mos* mRNA. *Proc Natil Acad Sci USA.* 93. 14602-14607.

Huarte, J., Belin, A., Vassali, S., and Strickland, S. (1987). Meiotic maturation of mouse oocyte triggers the translation and polyadenylation of dormant tissue-type plasminogen activator mRNA. *Genes Develop.* 1. 1201-1211.

Hwang, S.Y., Oh, B., Solter, D., Knowles, B., and Lee, J.S. (2001). Expression of genes involved in the mammalian meiosis during the transition from the egg to embryo. *Mol Rep Devel.* 59. 144-158.

Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics* 5. 150-163.

Fig. 3. Comparisons of the 3' end sequences between mammalian homologs of tPA (tissue-type plasminogen activator). Boxes with broken lines indicate the hexanucleotide nuclear polyadenylation signal. Previously known CPE sequences are marked by boxes with solid lines. An alternative CPE candidate identified in this study is double-underlined.

Lee, J.S., Choi, K.D., Lee, S.Y., Min, K.S., and Hwang, S.Y. (2006). Maternal contribution of cell death regulatory genes in early mouse and chick embryogenesis: an *in silico* analysis using EST database and homology search. *Korean J Genetics.* 28 107-117.

Liberfarb, M., Chu, T., Wreden, C., Theurkauf, W., Gergen, P., and Strickland, S. (1996). Mutations that perturb poly(A)-dependent maternal mRNA activation block the initiation of development. *Development.* 122 579-588.

Mendez, R., and Richter, J. (2001). Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol.* 2. 521-529.

Oh, B., Hwang, S.Y., Solter, D., and Knowles, B. (1997). *Spindlin*, a major maternal transcript expressed in the mouse during the transition from oocyte to embryo. *Development.* 124. 493-500.

Oh, B., Hwang, S.Y., McLaughlin, K., Solter, D., and Knowles, B. (2000). Timely translation during the mouse ooctye to embryo transition. *Development.* 127. 3795-3803.

Ohsugi, M., Hwang, S.Y., Butz, S., Knowles, B., Solter, D., and Kemler, R. (1996). Expression and cell membrane localization of catenins during mouse preimplantation development.

Dev Dynamic. 206. 391-402.

Paschoud, S., Dogar, A., Kruntz, C., Grisoni-Neupert, B., Richman, L., and Kuehn, L. (2006). Destabilization of interleukin-6 mRNA requires a putative RNA stem-loop structure, an AU-rich element, and the RNA-binding protein AUF1. *Mol Cell Biol.* 2006. 8228-8241.

Read, R., and Norbury, C. (2002). Roles for cytoplasmic polyadenylation in cell cycle regulation. *J Cell Biochem.* 87 258-265.

Richards, W., Carroll, P., Kinlock, R., Wassarman, P., and Strickland, S. (1993). Creating maternal effect mutations in transgenic mice: antisense inhibition of an oocyte gene product. *Dev Biol.* 160. 543-553.

Richter, J. (1999). Cytoplasmic polyadenylation in development and beyond. *Microbiol Mol Biol Rev.* 63. 446-456.

Seko, Y., Cole, S., Kasprzak, W., Shapiro, B., and Ragheb, J. (2006). The role of cytokine mRNA stability in the pathogenesis of autoimmune diseases. *Autoimmun Rev.* 5. 299-305.

Stebbins-Boaz, B., and Richter, J. (1997). Translational control during early development. *Crit Rev Eukayot Gene Expr.* 7. 73-94.