AU-rich elements (ARE) found in the U-rich region of Alu repeats at 3' untranslated regions.

Hyeong Jun An [1], Doheon Lee [1*], Kwang Hyung Lee [1] and Jonghwa Bhak [1,2,3,4*,#]

[1] Department of BioSystems, KAIST , Daejeon, Korea
[2] NGIC, KRIBB, Daejeon, Korea
[3] Object Interaction Technologies (OITEK), Inc. Daejeon, Korea
[4] BiO centre, Daejeon, Korea
JB's former English name is Jong H. Park.

*To whom correspondence should be addressed . E-mail: doheon@kaist.ac.kr, biopark@kaist.ac.kr

Abstract

A significant portion (about 8% in human genome) of mammalian mRNA sequences contains AU (Adenine and Uracil) rich elements or AREs at their 3' untranslated regions (UTR). These mRNA sequences are usually stable. ARE motifs are assorted into three classes. The importance of AREs in biology is that they make certain mRNA unstable. We analyzed the occurrences of AREs and Alu, and propose a possible mechanism on how human mRNA could acquire and keep A REs at its 3' UTR originated from Alu repeats .

Interspersed in the human genome, Alu repeats occupy 5% of the 3' UTR of mRNA sequences. Alu has poly-adenine (poly-A) regions at the end that lead to poly -thymine (poly-T) regions at the end of its complementary Alu. It has been discovered that AREs are present at the poly -T regions. In the all ARE's classes, 27-40% of ARE repeats were found in the poly -T region of Alu with mismatch allowed within 10% of ARE's length from the 3' UTRs of the NCBI's reference m RNA sequence database.

We report that Alu, which has been reported as a junk DNA element, is a source of AREs. We found that one third of AREs were derived from the poly -T regions of the complementary Alu.

Introduction

The half life of eukaryotic mRNA shows over 10 times difference [1,2]. This is partly because AU -rich elements (ARE), a *cis*-acting regulation sequence that are found in mRNA , have effect on the structure of mRNA and degradation factor proteins [3,4,33]. ARE is located at the end of mRNA and has an effect of making poly -adenine (Poly-A) shorter (deadenylation). When the poly-A becomes shorter, the interaction between PABP (Poly-A binding protein) that binds to poly-A at 3'of mRNA and cap structure at 5' becomes weak. Once this interaction i s broken, mRNA can be easily degraded by losing its stable structure [5 -9,34,38]. ARE research is important many ways

This work is supported by the Korean Systems Biology Research Grant and the Korea and BioGreen21 project.

especially as in proto-oncogenes that have AREs in their mRNA. Mutations in the proto-oncogenes convert them to oncogenes and can cause cancers. It is known that half life of the proto-oncogenes' mRNA is shorter than usual mRNA because of AREs [14-21,39]. Growth factors [22] and vascular endothelial factors [23] are also known to contain AREs.

AU-rich elements (ARE) is usually bound by mRNA degradation proteins. However, mRNA stabilization proteins can also bind ARE with no clear knowledge on the mechanism [13,40].

Table 1. The descriptions of AU-rich elements' classes

	Description
Class I	Scattered AUUUA in U stretches
Class II	(AUUU)*n*A, ($2 \leq n \leq 5$)
Class III	Non-AUUUA, U stretch

AU-rich elements (ARE) has a nonamer sequence such as UUAUUUAUU as a basic working element. Usually, higher number of AUUUA repeat causes faster degradation of mRNA. AREs can be classified into three types based on their structure (Table 1) [3,10-12,42].

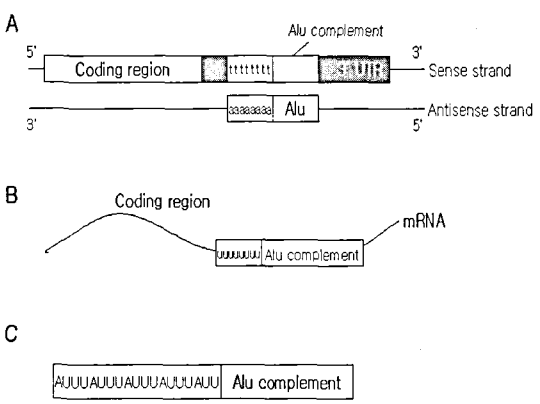
Class I has more than one AUUUA element at poly-uracil region. For example, UUUAUUUAUUUUUUAUUUAUUU contains two AUUUA in a poly-U sequence [3]. Class II contains repeated AUUUA element. For example, (AUUU)*n*A (n=5) or AUUUAUUUAUUUAUUUAUUUA are such [3,12]. Depending on the size of *n*, they can be subclassified. Class III is a non-AUUUA type ARE. They could contain poly-uracil. Also, non-AUUUA elements are found in this class [3,42]. Among the three, class II is best known [42]. The increased number of AUUU in Class II increases the efficiency in mRNA degradation [10].

Unlike ARE, Alu is regarded as a kind of junk DNA. Alu is transcribed by RNA polymerase III. The transcribed Alu mRNA are incorporated into DNA and the copy number rises over time [24,25]. Alu mRNA contains poly-A that resembles common mRNA and this poly-A plays an important role in the expansion of Alu in human genome [36, 37]. The length of Alu is about 300bp. It occupies 6-13% of human genome [31]. About 5% of human cDNA has Alu and the majority is found at 3'UTR [32]. Alu has three types depending on their age. They are Alu J, Alu S, and Alu Y. Alu J appeared 80 million years ago (mya). Alu S is younger and appeared around 60 mya. Alu Y is recent and is predicted to have appeared around 20,000 to 40,000 years ago. These three types can be subclassified. Alu S has as many as 12 subfamilies [26]. Alu Y is still active in the human genome [27-30]. When Alu mRNA is inserted, Target Site Duplication (TSD) is created. As a result, there are repeated sequence of 7-20bp bases at both sides of Alu [35]. Initially, Alu was regarded to have no function. However, recently, their function has been suggested. If situated at the upstream of genes, transcript factors (TF) can be bound resulting in an amplification of transcription [29]. Our group suggested the link between Alu at 3' UTR and pseudogenes [43]. However, there is no evidence for the association between ARE and 3'UTR Alu.

Here, we report a discovery of ARE that affects the half life of mRNA from the 3' UTR Alu sequences. As in figure 1A, when Alu is

inserted to the complementary strand of DNA coding region, the poly-A at the end of Alu becomes poly-thymine (T). If this gene is transcribed the mRNA will contain poly-U at 3'UTR. This poly-U is one of the ARE types (class III). We discovered ARE class II and I also have a link with Alu as well as ARE class III. We suggest this as a new function of 3'UTR Alu.

Figure 1. The schematic diagram of poly-thymine (poly-T) generation by Alu. (A) Alu contains poly-adenine (poly-A) region at the end. It is shown as 'aaaaaaaa'. The poly-A of Alu at anti-sense becomes poly-T (complement of poly-A) at the sense strand on DNA. It is shown as 'tttttttt'. (B) mRNA now contains poly-uracile (poly-U) region after the transcription of poly-T region. (C) AU-rich elements are found in this poly- U region in (B).



Materials and Methods

Human 3' untraslated region

We used RefSeq database from National Center for Biotechnology Information (NCBI) for human 3'UTR sequences [41]. We extracted 3' UTR of CDS (coding sequence) from all the annotated mRNA sequence (mRNA_Prot, 2004.9.13). The number of 3'UTR was 21,121 and the average length was 1.6kbp.

AU-rich elements (ARE) pattern searching

AREs were searched for in the 3'UTR for all three classes. We used (AUUU)*n*A (n=5, 21bp) pattern for ARE class II [12]. In table 1, class II is defined to have 2-5 repeats of AUUU. We used the longest repeat size only for achieving high accuracy. We used 30bp for Class I and III as the length definition for these classes are not accurate. They are usually referred to as long poly-U [3, 42]. 30bp is the minimum length observed in experiments [44].

Alu sequence detection

RepeatMasker program is used for finding Alu. It is a common program for finding repeats [45]. After finding Alu sequences using RepeatMasker at 3'UTR, for each Alu, we recorded the position information for the next step analysis.

Comparison between two search results

We compared the positions of 3'UTR Alu and ARE. If ARE is discovered within ARE, we used TSD (Target Site Duplication) information. For example, if Alu is found between 100- 400bp in 3'UTR and ARE between 90-120bp, we search for TSD up- and downstream of the region. If TSD is found, ARE is found to be inside of Alu region.

Statistical analysis of the search results

To validate the significance of the searches, we calculated the random chance of the ARE and Alu overlap.

Hypothesis

H0: ARE occurs in human 3'UTR independently from Alu.

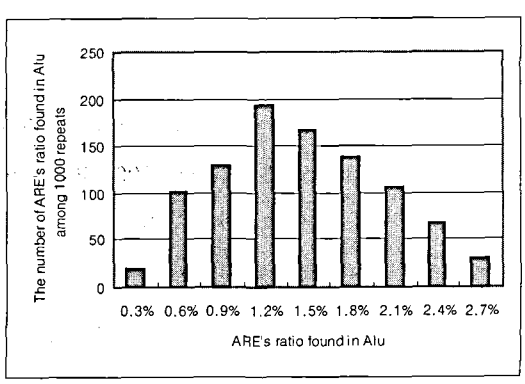
Validation

The average length of 3' UTR of 21,121 human sequences was 1,600 bp. Within the large sequence of 21,121 X 1,600 bp, we generated 1504 Alu (300 bp) and 329 ARE (30bp) following a uniform distribution. 1504 and 329 are the actual numbers of Alu and ARE found by our method. This random sequence generation is done 1,000 times with 95% significance threshold.

Test Result

The significance range at 5% error range was 0.3-2.7% (Figure 2) for randomness. Therefore, our search result of 27-40% is out of the likelihood for random overlaps of Alu and ARE. Therefore, the hypothesis H0 is rejected.

Figure 2. Empirical confidence interval of ARE's ratio found in Alu. X-axis was ARE's ratio found in Alu, and Y-axis was the number of ARE's ratio found in Alu among 1,000 repeats.



Results

About 30% of ARE were found in and probably originated from Alu.

The result from the method is shown in table 2. In ARE class I, out of the total 81 AREs, 23 of them were found to be in Alu region. Class II had 15 out of total 55 and the class III had 78 found by our method out of 193. Out of 329 ARE located in 3'UTR from the 21,121 mRNA, over 35% of them were within 1504 Alu sequences. As the random chance of ARE being located within Alu was from 0.3% to 2.7% (Statistical analysis in the Method) this result is highly significant and 3' UTR ARE are significantly associated with Alu.

Table 2. ARE ratio found in Alu by ARE' classes. (A) column is the number of ARE found in Alu. (B) column is all ARE found in the 3'UTR of 21,121 human mRNA sequences. (C) column is ARE ratio found in Alu (A/B).

	ARE found in Alu[A)]	all ARE in all 3'UTR[B)]	Ratio[C)]
Class I	23	81	28%
Class II	15	55	27%
Class III	78	193	40%
SUM	116	329	35%

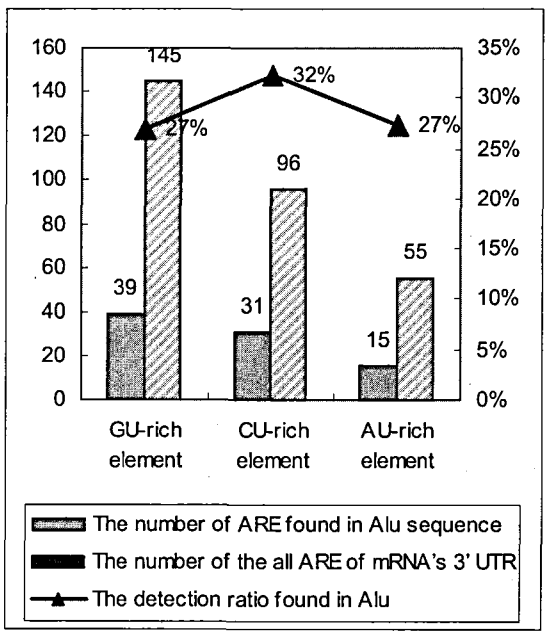
GU- or CU- rich elements were also found in Alu.

AUUUA repeats is not the only sequence pattern for ARE. Non-AUUUA AREs that contain repeats such as GUUUG can also be ARE. It was discovered that GUUUG repeats function as ARE in *c-jun*'s 3'UTR [42]. In addition to GUUUG repeats, CUUUC repeats also function as ARE even though mRNA stabilizing factors bind to it [40]. In general, theses two cases are classified into the class III ARE. Our search showed that GU or CU repeats were also found in Alu. We

found that GU- or CU- rich elements in place of adenine in the ARE class II were most likely to have occurred through adenine bases having been replaced with guanine or cytosine (i.e. (GUUU)*n*G (*n*=5) or (CUUU)*n*C (*n*=5)).

These G and C base types were also found to be associated with Alu in this study. Figure 3 shows the number and ratio of GU- and CU-rich elements associated with Alu. Out of 145 total GU-rich elements, 39 were found with Alu. CU-rich elements had 31 out of 96. The overall ratio of occurrence with Alu sequence is in between 27-32% which is similar to other AREs (Table 2).

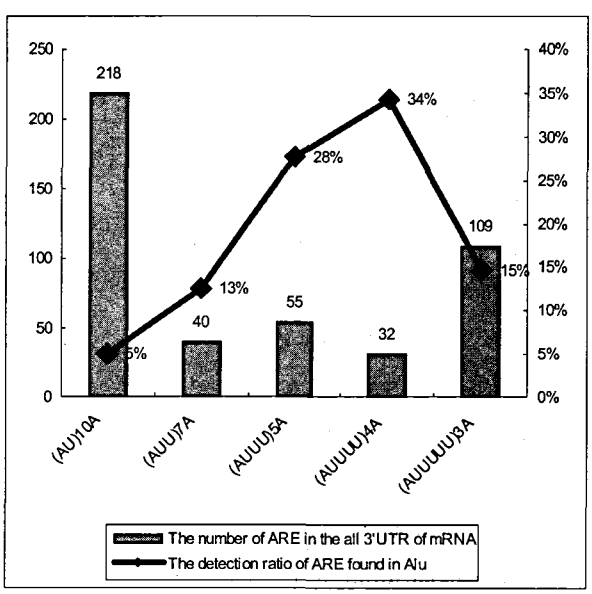
Figure 3. GU- or CU- rich elements found in Alu. X-axis was each GU-, CU-, and AU-rich elements. AU-rich elements were class II type in this figure. Two Y-axes were the number of each element (left) and ARE's ratio found in between Alu and all mRNA.



AUUU or AUUUU repeats were found more frequently in Alu than other repeats such as AU or AUUUUU repeats.

We tested whether insertion/substitution at every 4bp (AUUU repeats) was more frequently found than other base pairs (AU, AUU, AUUUU or AUUUUU repeats) in 3'UTR Alu. The result can be dependent on the full length of repeats. For example, (AUUUUU)3A occurred 109 times, and (AUUUUU)4A was 31 times in all the 3'UTR sequences. However, the two ratios found in Alu were similar; 16 out of 109 (AUUUUU)3A (15%) and 5 out of 31 (AUUUUU)4A (16%) were found in Alu. We tested all the repeats of AU, AUU, AUUU, AUUUU and AUUUUU that are closest to 21bp. Therefore, each repeat element had difference of 0-2bp.

Figure 4. The ARE found in Alu by changing the number of uracil repeat between the two flanking adenines of AUUUA motif. The full length of each repeat had the minimum size of 21bp.



Interestingly, among the repeat elements, we found that ARE class II (AUUU)5A and

(AUUUU)4A were highly associated with Alu while other repeats showed less than 15% occurrences with Alu.

Discussion

The function of Alu has not been clearly known and regarded as the vestige of molecular evolution. Recently, it has been reported that it affects the transcription and splicing at the upstream gene and intron regions. However, no critical biological function is known for the Alu in downstream 3'UTR.

In this study, we discovered for the first time that Alu inserted at the 3' UTR is associated and probably the source of ARE that regulates the degradation of mRNA. We discovered that around one third of ARE were statistically significantly associated with and probably originated from Alu. We found that adenines were inserted or substituted regularly in poly-U (Figure 4). Out of the institutional/substitutional range of 4-5bp (the repeats of AUUU or AUUUU), the occurrence of ARE in Alu region decreases in the repeats of AU, AUU, or AUUUUU. We suggest that there is a molecular mechanism that determines the regular interval of 4-5bp. We also examined non-adenine replacement of uracil (U) and they had the same regular 3-5 bp replacement (Figure 3). The order of bases is guanine > cytosine > adenine. Regardless of the base types, around 30% of such sequence patterns were found in Alu region indicating the Alu origin or association with Alu.

In conclusion, Alu's poly-U alone can affect the degradation of mRNA, regular insertion/substitution of adenine in every 3bp seemed to have raised the efficiency of ARE. As other bases showed similar patterns, we suggest that the type of base is not the most critical element in mRNA degradation. This study contributes significantly to the regulation of mRNA degradation by ARE and suggest a new function of Alu in human and possibly other high mammalian genomes.

Acknowledgements

We thank Dr. Kim, Ho at SNU for his statistical expertise. HJA, DL, and KHL are supported by Korean Systems Biology Research Grant (M1-0309-02-0002), JHB is supported by BioGreen21 and IMT-2000-C4-3 from ministry of information and communication of Korea. We thank CHUNG Moon Soul Center for BioInformation and BioElectronics and IBM SUR program.

References

1. Beelman CA, Parker R: Degradation of mRNA in eukaryotes. Cell 1995, 81: 179-183

2. Tucker M, Parker R: Mechanisms and control of mRNA decapping in Saccharomyces cerevisiae. Annu. Rev. Biochem. 2000, 69: 571-595

3. Chen CY, Shyu AB: AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. 1995, 20:465-470

4. Ambro H, Parker R: Messenger RNA degradation: Beginning at the end. Curr. Biol. 2002, 12:R285-R287

5. Muhlrad D, Decker CJ, Parker R: Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'->3' digestion of the transcript. Genes Dev. 1994, 8:855-866

6. Jacobs Anderson JS and Parker R: The 3' to 5'

degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J. 1998, 17:1497-1506

7. Chen CY, Gherzi R. Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M. Karin M: AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell 2001, 107:451-464

8. Wang Z, Kiledjian M: Functional link between the mammalian exosome and mRNA decapping. Cell 2001, 107: 751-762

9. Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijn GJ, Lutz CS, Wilusz J: The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. EMBO J. 2002, 21:165-174

10. Zubiaga AM, Belasco JG, Greenberg ME: The nonamer UUAUUUAUU is the key AU-Rich sequence motif that mediates mRNA degradation. Mol. Cell. Biol. 1995, 15:2219-2230

11. Akashi M, Shaw G, Hachiya M, Elstner E, Suzuki G, Koeffler P: Number and location of AUUUA motifs: role in regulating transiently expressed RNAs. Blood 1994, 83: 3182-3187

12. Bakheet T, Frevel M, Williams BR, Greer W, Khabar KS: ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. Nucleic. Acids. Res. 2001, 29:246-54

13. Park-Lee S, Kim S and Laird-Offringa IA: Characterization of the Interaction between Neuronal RNA-binding Protein HuD and AU-rich RNA. J. Biol. Chem. 2003, 278:39801-39808

14. Chen CY, Chen TM, Shyu AB: Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. Mol. Cell. Biol. 1994, 14:416-426

15. Reeves R, Magnuson NS: Mechanisms regulating transient expression of mammalian cytokine genes and cellular oncogenes. Prog. Nucleic Acid Res. Mol. Biol. 1990, 38:241-282

16. Brewer G: An A + U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro. Mol. Cell. Biol. 1991, 11:2460-2466

17. Wingett D, Reeves R, Magnuson NS: Stability changes in pim-1 protooncogene mRNA after mitogen stimulation of normal lymphocytes. J. Immunol. 1991, 147:3653-3659

18. Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A: Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl Acad. Sci USA. 1986, 93: 1670-1674

19. Gorospe M, Baglioni C: Degradation of unstable interleukin-1 alpha mRNA in a rabbit reticulocyte cell-free system. Localization of an instability determinant to a cluster of AUUUA motifs. J. Biol. Chem. 1994, 269:11845-11851

20. Peppel K, Vinci JM, Baglioni C: The AU-rich sequences in the 3'-untranslated region mediate the increased turnover of interferon mRNA induced by glucocorticoids. J. Exp. Med. 1991, 173:349-355

21. Gillis P, Malter JS: The adenosine-uridine binding factor recognizes the AU-rich elements of cytokine, lymphokine, and oncogene mRNAs. J. Biol. Chem. 1991, 266:3172-3177

22. Sirenko OI, Lofquist AK, DeMaria CT, Morris JS, Brewer G, Haskill JS: Adhesion-dependent regulaton of an A+U rich element-binding activity associated with AUF1. 1997, 17:3898-3906

23. Pages G, Berra E, Milanini J, Levy AP, Pouyssegur J: Stress-activated protein kinases (JNA and p38/HOG) are essential for vascular endothelial growth factor mRNA stability. J. Biol. Chem. 2000, 275:26484-26491

24. Rogers J: Retroposons defines. Nature 1998, 301:460.

25. Deininger PL, Batzer MA: Alu repeats and human disease. Mol. Genet. Met. 1999, 67:183-193

26. Shen M, Batzer MA, Deininger PL: Evolution of the mater Alu gene(s). J. Mol. Evol. 1991, 33:311-320.

27. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E: Standardized Nomenclature for Alu Repeats. J. Mol. Evol. 1996, 42:3-6.

28. Bailely AD, Shen, CK: Sequential Insertion of Alu Family Repeats into Specific Genomic Sites of Higher Primates. Proc. Natl. Acad. Sci. USA 1994, 90:7205-7209

29. Britten RJ: Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. Proc. Natl. Acad. Sci. USA 1994, 91:6148-6150

30. Kapitonov V, Jurka J: The age of Alu subfamilies J. Mol. Evol. 1996, 42:59-65

31. Boeke JD: LINE and Alus the polyA connection. Nature Genet. 1997, 16:6-7

32. Yulug IG, Yulung A, Fisher EM: The frequency and position of Alu repeats in cDNAs, as determined by database searching. Genomics 1995, 27:544-548

33. Vasudevan S, Peltz SW: Regulated ARE-mediated mRNA decay in Saccharomyces cerevisiae. Mol. Cell 2001, 7:1191-1200

34. Hoof AV, Parker R: The exosome: a proteasome for RNA? Cell 1999, 99:347-350

35. Kazazian HH Jr: Mobile elements: Drivers of genome evolution. Science 2004, 303:1626-1632

36. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, and Deininger PL: Active Alu element "A-tails": size does matter. Genome R. 2002, 12:1333-1344

37. Ostertag EM, Kazazian Jr HH: Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. 2001, 35:501-538

38. Mukherjee D, Gao M, O'Connor JP, Raijmakers R, Pruijin G, Lutz CS, Wilusz J: The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. EMBO J. 2001, 21:165-174

39. Wilson T, Treisman R: Revomal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. Nature 1988, 336:396-399

40. Wein G, Roessler M, Klug R, Herget T: The 3'-UTR of the mRNA coding for the major protein kinase C substrate MARCKS contains a novel CU rich element interacting with the mRNA stabilizing factors HuD and HuR. Eur. J. Biochem. 2003, 270:350-365

41. Pruitt KD, Tatusova T, Maglott DR: NCBI Reference Sequence Project: update and current status. Nucleic. Acids. Res. 2003, 31:34-37

42. Peng S S-Y, Chen C-Y A., Shyu A-B:

Functional characterization of a non-AUUUA AU-Rich element from the c-jun proto-oncogene mRNA: Evidence for a novel class of AU-Rich elements. Mol. Cell. Biol. 1996, 16:1490-1499

43. An HJ, Na D, Lee D, Lee KH, Bhak JH: High correlation between Alu elements and the conversion of 3'UTR of mRNAs processed pseudogenes. Genomics & Informatics 2004, 2:94-100

44. Stoecklin G, Stoeckle P, Lu M, Muehlemann O, Moroni C: Cellular mutants define a common mRNA degradation pathway targeting cytokine AU-rich elements. RNA 2001, 7:1578-1588

45. Smit AFA, Hubley R, Green P: RepeatMasker Open-3.0.1996-2004, http://www.repeatmasker.org