

This Provisional PDF corresponds to the article as it appeared upon acceptance. The fully-formatted PDF version will become available shortly after the date of publication, from the URL listed below.

The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3'untranslated regions.

BMC Genomics 2004, 5:97 doi:10.1186/1471-2164-5-97

Hyeong Jun An (hjan@bisl.kaist.ac.kr)
Doheon Lee (doheon@kaist.ac.kr)
Kwang Hyung Lee (khlee@bisl.kaist.ac.kr)
Jonghwa Bhak (biopark@kaist.ac.kr)

ISSN 1471-2164

Article type Research article

Submission date 4 Aug 2004

Acceptance date 21 Dec 2004

Publication date 21 Dec 2004

Article URL <http://www.biomedcentral.com/1471-2164/5/97>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

The association of Alu repeats with the generation of potential AU-rich elements (ARE) at 3' untranslated regions.

Hyeong Jun An¹, Doheon Lee^{1§}, Kwang Hyung Lee¹, and Jonghwa Bhak^{1,2,3,4§}

¹BioSystems Dept., Korean Advanced Institute of Science and Technology (KAIST)
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea

²NGIC, KRIBB, Daejeon, Korea

³BiO institute, Daejeon, Korea

⁴OITEK (Inc), Daejeon, Korea

[§]Corresponding authors: Doheon Lee and Jonghwa Bhak

Email addresses:

HJA: hjan@bisl.kaist.ac.kr

DL: doheon@kaist.ac.kr

KHL: khlee@bisl.kaist.ac.kr

JB: biopark@kaist.ac.kr

Abstract

Background

A significant portion (about 8% in the 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. However, an increasing number of observations have been made of unstable species, possibly depending on certain elements such as Alu repeats. ARE motifs are repeats of the tetramer AUUU and a monomer A at the end of the repeats ((AUUU)_nA). The importance of AREs in biology is that they make certain mRNA unstable. Proto-oncogene, such as *c-fos*, *c-myc*, and *c-jun* in humans, are associated with AREs. Although it has been known that the increased number of ARE motifs caused the decrease of the half-life of mRNA containing ARE repeats, the exact mechanism is as of yet unknown. We analyzed the occurrences of AREs and Alu and propose a possible mechanism for how human mRNA could acquire and keep AREs at its 3' UTR originating from Alu repeats.

Results

Interspersed in the human genome, Alu repeats occupy 5% of the 3' UTR of mRNA sequences. Alu has poly-adenine (poly-A) regions at its end, which lead to poly-thymine (poly-T) regions at the end of its complementary Alu. It has been found that AREs are present at the poly-T regions. From the 3' UTR of the NCBI's reference mRNA sequence database, we found nearly 40% (38.5%) of ARE (Class I) were associated with Alu sequences (Table 1) within one mismatch allowance in ARE sequences. Other ARE classes had statistically significant associations as well. This is far from a random occurrence given their limited quantity. At each ARE class, random distribution was simulated 1,000 times, and it was shown that there is a special relationship between ARE patterns and the Alu repeats.

Conclusion

AREs are mediating sequence elements affecting the stabilization or degradation of mRNA at the 3' untranslated regions. However, AREs' mechanism and origins are unknown. We report that Alu is a source of ARE. We found that half of the longest AREs were derived from the poly-T regions of the complementary Alu.

Background

Varying more than ten-fold, messenger RNA degradation is essential for the regulation of gene expression [1, 2]. Differential mRNA decay rates were determined by specific *cis*-acting sequences within mRNA. For example, the mRNA sequences of yeast, many mammals, and other eukaryotes contain AU-rich elements or AREs at their 3' untranslated regions (UTR) [3, 4]. For example, in yeast, AREs stimulated the shortening of poly adenine (poly A), and two kinds of degradation pathways followed. One is 5'-to-3' exonuclease access by removal of the 5' cap structure. The other is 3'-to-5' digestion by a complex of exonucleases called exosome [5, 6]. Genes required for these steps have been identified in yeast and were found to be conserved among eukaryotes. Although the mechanisms of AREs enhanced mRNA degradation are unknown, several groups provided evidence that 3'-to-5' degradation by the exosome may be the major pathway of decay for at least some mammalian mRNAs, including

ARE-containing mRNA sequences [7, 8, 9]. The length of AREs also affected the half-life of mRNA. The nonamer UUAUUUAUU is a typical ARE, and the simple repeats, (AUUU)_nA motif, is the well-known pattern of AREs. It has been shown that the number of ARE motifs correlated with the turnover of ARE-mRNAs such as GM-CSF [10, 11]. Because of this, AREs are usually classified according to the number of the repeats [12].

It is known that the stabilization factor, such as HuD, is able to bind to AREs [13] and most AREs seem to function as destabilizing factors. The overall importance of AREs in biology is that they can make certain critical gene products unstable. They include proto-oncogenes such as c-fos [14], c-myc [15], c-myc [16], and Pim-1 [17]. Another class of ARE-associated genes are immune response genes such as interferon [15, 18] and interleukin [15, 19, 20, 21]. Growth factors, such as Gro- α [22] and the vascular endothelial factor [23] in humans, are also known to be associated with AREs.

AREs consist of a great number of thymine (or uracil) and a few adenines. Alu repeats can be a source of poly-T regions in mRNA. Therefore, there is a possible link between ARE and Alu repeats.

Alu repeats are sequences of approximately 300 nucleotides (nt) transcribed by RNA polymerase III. The Alu region is then reverse-transcribed and inserted into a new location in the genome [24]. It can reach a copy number in excess of 500,000 in the human genome [25]. Alu repeats were thought to be inserted very early in primate evolution, approximately 65 million years ago (mya). Alu amplification appears to have reached a maximum rate between 35 and 60 mya, and is currently amplifying at only 1% of the maximum rate [26]. Statistical analyses have identified key diagnostic nucleotide positions in Alu sequences that define 12 subfamilies. J class is the oldest one, S class is intermediate, and Y class is the newest. The majority of Alu retrotranspositions were completed at least 30 mya when the Alu-Sx subfamily, which accounts for half of all human Alu sequences, and the Alu-Sp and Alu-Sq subfamilies became unable to replicate [27, 28, 29, 30]. Alu repeats account for 6-13% of the human genome [31] and were identified in 5% of 1,616 human full-length cDNA. Of the 5%, 82% were found in the 3' UTR, while 14% were located in the 5' UTR, and very rarely in the coding region [32]. The common role of Alu at 3' UTR has not been reported, although there is one specific case that the chemical, PMA, can bind to Alu at 3' UTR and increased mRNA half-life [43].

We investigated the link between Alu sequence and the potential AREs (that have not been experimentally verified but contain ARE sequence patterns), and suggest that the complementary poly-adenine regions of Alu is one of the sources of AREs at the 3' UTR of mRNA. Figure 1A shows that the poly-adenine regions of Alu contained in the anti-sense strand on DNA complemented the poly-thymine regions in the sense strand; therefore, the poly-thymine regions on DNA transcribed the poly-uracil regions on mRNA (Figure 1B). We propose a mechanism on how Alu has been converted to AREs gradually. When adenine was inserted at a regular interval in the poly-T(U) regions, it eventually led to the generation of potential AREs. It is not clear why such a regular insertion occurs, but the phenomenon has also been found in other ARE-like sequences. Figure 1C shows transcribed ARE on mRNA [33, 34].

Results

The results from the method are shown in Figure 2. In the ARE class I, marked as (AUUU)₅A pattern in Table 1, 26 AREs were found in all 21,121 mRNA 3' UTR.

38.5% of 26 AREs included in the class I, were detected in Alu sequences at 3' UTR. When we did a simulation test for the 26 AREs and 1,504 Alu sequences by 1,000 times, with a 95% confidence interval (C.I.) threshold, it was statistically significant (see the statistical analysis of the search results in the Methods section). In other words, 38.5% occurrences were out of the likelihood for random overlaps of Alu and ARE patterns in the human genome. In the ARE class II (Table 1, (AUUU)4A pattern), 41 were found in all 3' UTR, and 7 were detected in Alu sequences among them (17.1%). The simulation results showed the 17.1% was less than the maximum random range of 7.3%. Therefore, class II data also showed a significance between ARE patterns and Alu. In class III (Table 1), 94 AREs were discovered from all 3' UTR. 15 out of 94 AREs were located in Alu sequences (16.0%). 16% was also statistically significant with the given sample size. In classes IV and V, 5% and 6.1% of ARE were found in Alu, respectively. These results were still out of the random chance distribution, although they were relatively less significant than the previous classes. In class VI, only 85 out of 8,649 AREs were detected in Alu (1%), and it is an insignificant hypothesis that the class VI pattern is associated with Alu sequences.

Discussion

The possible mechanism of how AREs originated from Alu is as follows: Alu is a special sequence that contains a poly-adenine (poly-A) region at its end. The poly-A region plays an important role in the retroposition mechanism of Alu [35]. It is known that the products of LINE (L1) transposon bind the poly-A of Alu. This enables Alu to retroposition [36, 37]. When Alu with poly-A are inserted as above, it is in the double helix form with the complementary poly-T. Therefore, the poly-T regions produce poly-uracil (poly-U) regions in mRNA when transcribed (Figure 1). We hypothesized that the poly-U regions generated from the Alu are the source of AREs after either random or directed mutation.

With this hypothesis, we suggest a new role for Alu was involved in the 3' UTR. It is well known that Alu affected gene expression at the 5' of genes and alternative splicing at the intron region [38, 39]. However, no Alu role at the 3' UTR has been suggested yet. We could have applied the same test to Alu at 5' UTR region, but there were too few data sources [32].

Conclusion

AREs are mediating sequences that affect the stabilization or degradation of biologically important genes' mRNA. However, their origin in evolution has not been clear. This report presents a hypothesis and statistical evidence that Alu was one of the sources of ARE generation or origin. A possible mechanism of ARE generation from Alu via retroposition and regular pattern mutation is suggested.

Methods

Human 3' UTR sequences

We used the RefSeq database from the 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 sequences (mRNA_Proc, 2004.9.13). The number of 3' UTR was 21,121 and the average length was 996bp. We used the Biojava package [42] to extract only 3' UTR with Genbank's feature information. The number of 3' UTR was 21,121 and the average length was 996bp.

Alu sequence and AU-rich element (ARE) pattern detection

AREs were searched for in the all 3' UTR (Table 1). An in-house java program was used to search for these AREs. While the number of AUUUA repeats decreased, the T flank region increased to 21bp. Each ARE was allowed within one base mismatch. This is a stricter mismatch criterion than the one of AU-rich elements database (ARED) (the ARED trained experimental ARE data allow 10% of ARE length mismatch [24]). The RepeatMasker program was used for finding Alu. It is a program for finding repeat sequences [25]. After finding Alu sequences using RepeatMasker at 3'UTR, for each Alu, we recorded the position information (RefSeq ID, start and end position) for the next step analysis.

Comparison between two search results

We compared the positions of 3' UTR Alu and ARE sequences. If an ARE was discovered within an Alu sequence, this ARE was regarded found in 3' UTR Alu. For example, if an Alu was found between 100-400bp and an ARE was found between 99-129bp, this ARE was in 3' UTR Alu in the same 3' UTR. If less than 50% of an ARE length was discovered in an Alu, we further check if there is 7bp TSD (Target Site Duplication) between the Alu's end and the ARE's end [4]. For example, if an Alu is between 100- 400bp and an ARE between 80-110bp, about 10bp (33%) of the ARE belongs to the Alu. In this case, we check if there is 7bp TSD between upstream region from 80bp and downstream from 400 bp.

Statistical analysis of the search results

To validate the significance of the searches, we calculated the random chance of the ARE and Alu sequence overlap at each class (Table 1).

Hypothesis

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

Random sequence generation for statistical validation

The average length of 3' UTR of 21,121 human sequences was 996bp. Within the long theoretical sequence of 21,121 X 996bp, we generated 1,504 Alu (300bp) and ARE sequences (21-13bp). For example, 1,504 Alu and 26 (21bp) AREs in ARE class I (Table 1) were generated following a uniform distribution as a control set. 1,504 and the number of AREs for ARE classes were the actual numbers of Alu and AREs found by our method. This random sequence generation was done 1,000 times with a 95% significance threshold.

Test Results

In the ARE class I (Table 1), the significance range at a 5% error range was 0.0-11.5% (Figure 2) for the random chance of association between ARE patterns and Alu sequences. The results in other ARE classes are also shown in Figure 2. Our

result of a 38.5% - 6.1% overlap between AREs and Alu, depending on ARE classes, was statistically significant. Therefore, hypothesis H0 was rejected.

Authors' contributions

HJA conceived of this study, carried out the tests, and drafted the manuscript. JB participated in the design of the study and drafted the manuscript. KWL and DL amended and improved the design of the study. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by Korea Research Foundation Grant (KRF-2003-041-D20490). JB is supported by IMT-2000-C4-3 grant of ministry of information and communication of Korea and BioGreen21 project of Korea. We would like to thank CHUNG Moon Soul Center for BioInformation and BioElectronics, and the IBM SUR program for providing research and computing facilities. We thank Maryana Huston for editing this manuscript and Dr. Kim, Ho at SNU for his statistical expertise.

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. Muhlrud 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**:246-254.
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 proto-oncogene 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 U S A 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 regulation of an A+U rich element-binding activity associated with AUF1.** Mol Cell Biol 1997, **17**:3898-3906.
23. Pages G, Berra E, Milanini J, Levy AP, Pouyssegur J: **Stress-activated protein kinases (JNK 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 Metab 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 U S A 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 U S A 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 Res 2002, **12**:1333-1344.
37. Ostertag EM, Kazazian Jr HH: **Biology of mammalian L1 retrotransposons.** Annu Rev Genet 2001, **35**:501-538.
38. Britten RJ, Davidson EH: **Gene regulation for higher cells: a theory.** Science 1969, **165**:349-357.
39. Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A: **Widespread RNA editing of embedded Alu elements in the human transcriptome.** Genome Res 2004, **14**:1719-1725.
40. **Refseq sequence database**
[ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/mRNA_Prot/]
41. **Biojava package**
[http://www.biojava.org]
42. **RepeatMasker Open-3.0.1996-2004**
[http://www.repeatmasker.org]
43. Wilson GM, Vasa MZ, Deeley RG: **Stabilization and cytoskeletal-association of LDL receptor mRNA are mediated by distinct domains in its 3' untranslated region.** J Lipid Res 1998, **39**:1025-1032.

Figures

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 sense strand on DNA. It is shown as 'ttttttt'.
- (B) mRNA now contains a 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).

Figure 2 - ARE found in Alu at each class (Table 1).

The numbers of ARE found in all 3' UTR, the number of ARE found in the Alu sequence, the ratio between them, and the randomly simulated results among 1,000 times at each ARE class (Table 1). Only the maximum possible ratios of the randomly simulated range at 95% confidence interval (C.I.) were shown. X-axis is for ARE patterns in all the classes. The left Y-axis is for the number of AREs, and the right Y-axis is for the overlap ratios.

Table

Table 1 - Defined ARE classes. (Symbol marks are used in this study instead of full sequences.)

	Symbol	ARE sequence
Class I	(AUUU)5A	AUUUAUUUAUUUAUUUAUUUA
Class II	(AUUU)4A	AUUUAUUUAUUUAUUUA
Class III	U(AUUU)3AU	UAUUUAUUUAUUUAU
Class IV	UU(AUUU)2AUU	UUAUUUAUUUAUU
Class V	U4AUUU4	UUUUUUUUUUUUUUUUUUUU
Class VI	W3UAUUUAUW3	WWWUAUUUAWWW

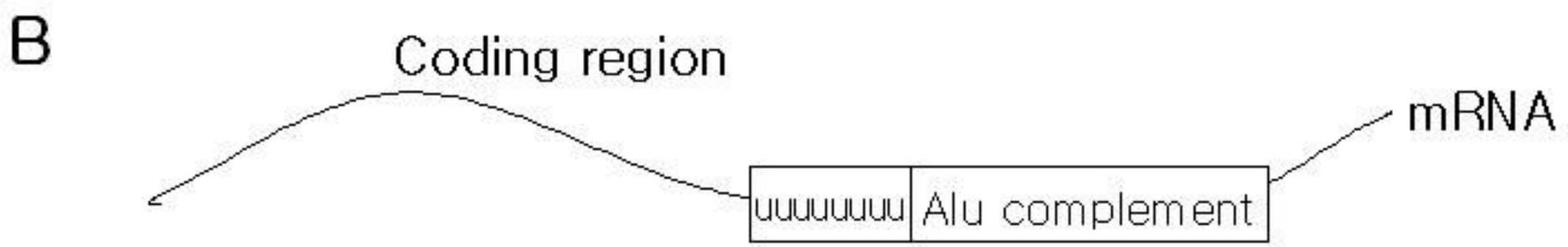
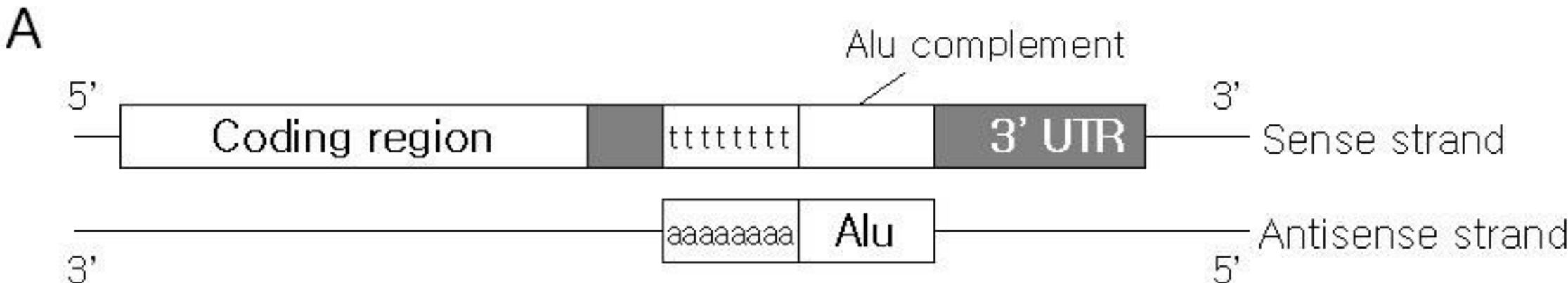


Figure 1

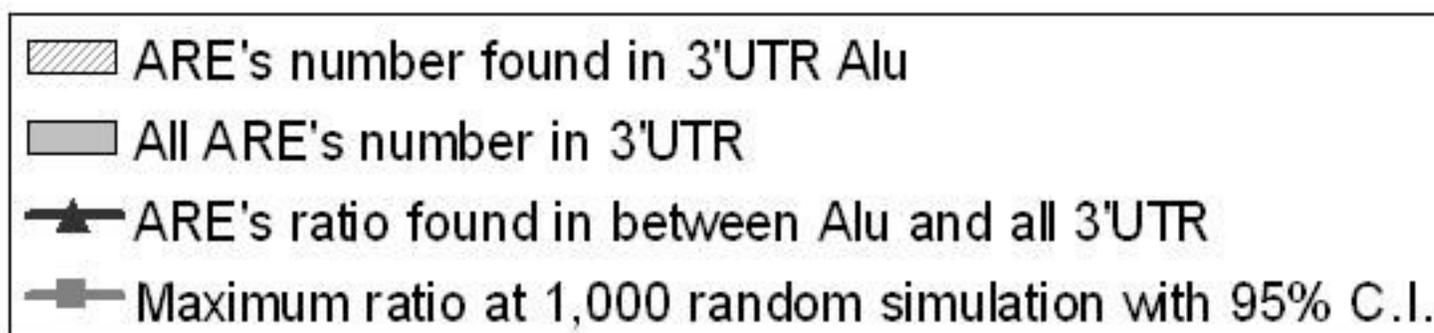
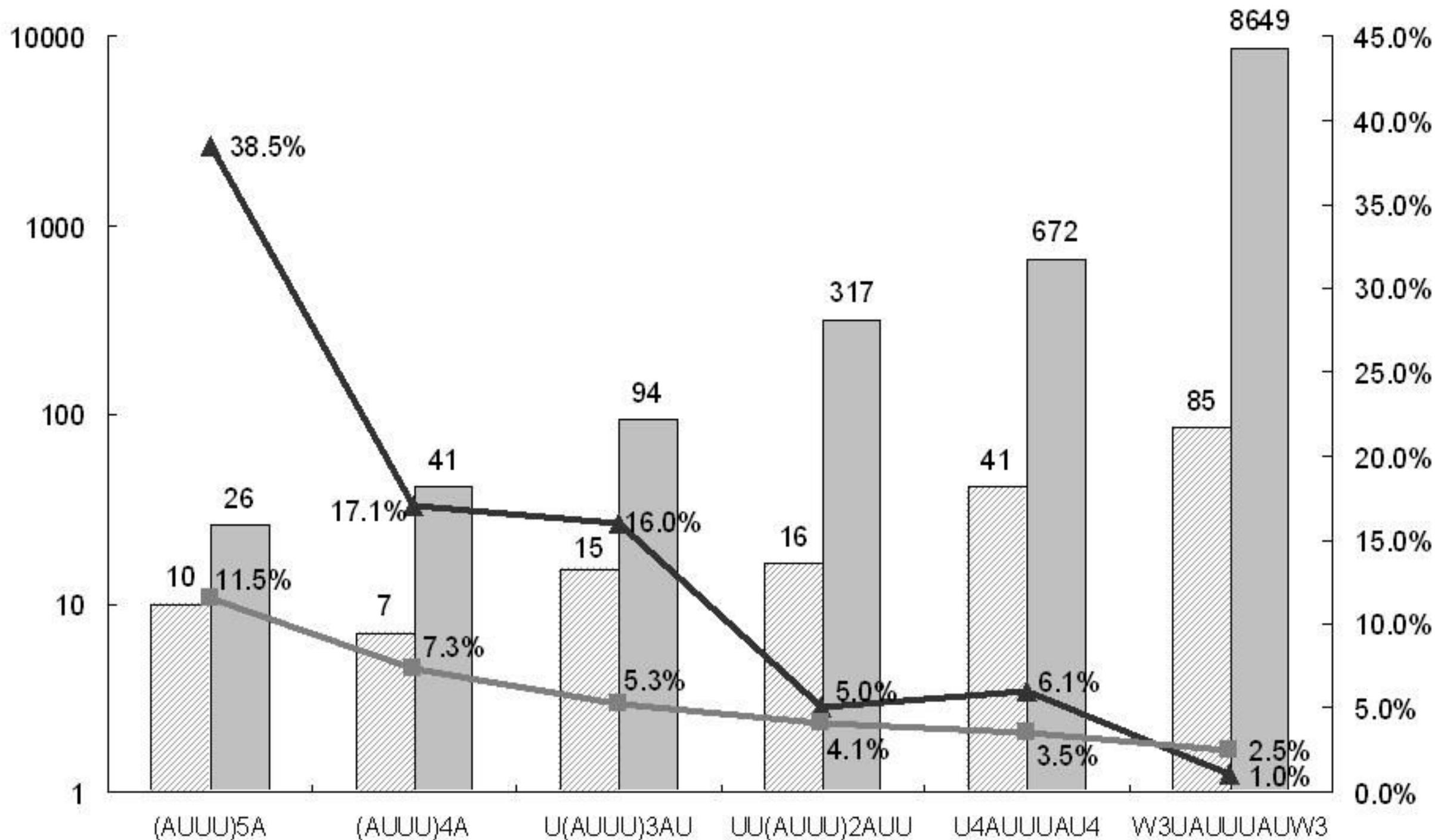


Figure 2