Chapter-V

Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells
5.1 Introduction

Actin is an abundantly expressed globular structural protein found in eukaryotes, involved in muscle contraction, cell mobility, cytokinesis, organelle movement and maintenance of cell shape. In lower organisms like yeast, actin is coded for by a single gene whereas different isoforms of actin perform different functions in higher organisms like mammals. The expression of cytoplasmic beta-actin gene is generally believed to be constitutive and ubiquitous and it is well known that it carries out housekeeping functions in the eukaryotic cell. However, recently, actin has been shown to play diverse roles besides its housekeeping functions. Our own previous studies have shown differential expression of the Actb gene in cellular model of SCA17 (Ghosh T et al., 2007, Chapter-II in this thesis). Presence of actin in the nucleus has been demonstrated and functionally, it has been shown to interact with the transcriptional machinery, suggesting a role in transcriptional regulation (Miralles F and Visa N, 2006; Bettinger BT et al., 2004; Hu P et al., 2004; Hofmann WA et al., 2004; Grummt I, 2006). In several cell types, including chicken fibroblasts and mammalian neuronal cells actin mRNA is known to be selectively localized to the cell periphery. The high conservation of the 3’ untranslated region of actin in vertebrates led to the suggestion that they may contain regulatory sequences (Ponte P et al., 1984). Subsequently, it has been demonstrated that a 54 nucleotide element called the zipcode, immediately following the stop codon is necessary for the localization of actin mRNA (Kislauskis EH et al., 1994). Actin is also known to be deadenylated during early developmental stages along with many other transcripts (Bachvarova R et al., 1985; Paynton BV et al., 1988). Besides its role in localization, the 3’ UTR of the actin mRNA has not been extensively studied and barring EST based evidence, there exists no experimentally determined polyadenylation site in the actin transcript.

Polyadenylation of eukaryotic transcripts is a two step enzymatically driven process, wherein mRNA is cleaved at a specific site and subsequently, the action of polyA polymerase, in the presence of a large basal polyadenylation machinery, adds adenosine nucleotides to the mRNA (Wahle E and Keller W, 1992; Keller W, 1995; Manley JL, 1995; Proudfoot N, 1996; Zhao J et al, 1999; Edmonds M, 2002; Gilmartin GM, 2005). Several factors like the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF) and cleavage factors...
(CFs I and II) perform specific functions within the polyadenylation machinery. Polyadenylation is important in conferring transcript stability (Ross J, 1995; Ford LP et al., 1997) and translational efficiency (Jackson RJ and Standart N, 1990; Sachs A and Wahle E, 1993; Sheets MD et al., 1994; Curtis D et al., 1995; Preiss T et al., 1998). The polyA tail is also believed to target the RNA for nuclear export (Eckner R et al., 1991; Huang Y and Carmichael GG, 1996). Although the enzymatic process of polyadenylation is well understood, the sequences in the mRNA that specify the site of cleavage and extent of polyadenylation are not fully characterized. In a majority of cases, cis-acting elements surrounding the polyadenylation site act together to specify the cleavage site. The sequence motif AAUAAA referred to as polyadenylation signal (PAS) that occurs about 10-30 nt (Proudfoot N, 1991; Colgan DF and Manley JL, 1997) upstream of the polyadenylation site in majority of human transcripts (Beaudoing E et al., 2000; Tian B et al., 2005), binds to the 160kDa subunit of CPSF. However, variants of the AAUAAA motif can effectively drive polyadenylation (Beaudoing E et al., 2000; Tian B et al., 2005). Besides the PAS element, G/U or U rich sequences downstream to the PAS contribute to the polyadenylation site (Wahle E and Ruegsegger U, 1999; Tabaska JE and Zhang MQ, 1999; Legendre M and Gautheret D, 2003) by binding to the multi-protien CstF.

Recently, bioinformatics analyses of the occurrence of polyadenylation sites in the human and mouse genomes have shown a surprisingly large number of alternative polyadenylation sites (Yan J and Marr TG, 2005; Tian B et al., 2005). Large scale mapping of EST sequences to the human and mouse genome showed that alternative polyadenylation is widespread in the mouse and human genomes. The hexamer motifs present upstream of the polyadenylation sites identified by EST analysis vary significantly from the canonical motif, AAUAAA. Ten to thirteen variants of the PAS element account for the polyadenylation sites identified over the whole genome. The PAS hexamers AAUAAA and AUUAAA together account for nearly 69% of all polyadenylation sites in the human genome (Tian B et al., 2005). The length of the 3’UTR, defined by the distance between the stop codon and the proximal polyA site was highly variable with a median length of 324 in humans and 385 nt in mouse (Tian B et al., 2005). The frequent occurrence of polyadenylation sites within internal exons and introns of genes implies that the use of alternative polyadenylation sites contributes significantly to the complexity of the eukaryotic transcriptome. When the
Alternative sites occur within protein coding regions of the transcript, it can result in functionally different products belonging to the same gene family, as shown in the case of the Lamin genes (Yan J and Marr TG, 2005). Transcripts coding for the same protein but containing different 3’UTRs may differ in stability, localization, translational efficiency and targeting by small regulatory RNAs. Such cases called tandem polyA sites (Edwalds-Gilbert G et al., 1997) have been shown in several genes and implicated in differential expression of transcripts in tissues. For instance, the transcripts from the huntingtin gene, that causes the Huntington’s neurodegenerative disorder, are preferentially polyadenylated at the distal site in the brain, but truncated at the proximal polyadenylation site in other tissues (Lin B et al., 1993). Testis, is one of the tissues in which frequent use of alternative tandem polyadenylation site has been observed (Edwalds-Gilbert G et al., 1997). At least two groups have independently reported the biased use of alternative polyadenylation sites in different tissues. By mapping ESTs from libraries derived from different tissues, Beaudoin et al (Beaudoin E and Gautheret D, 2001) showed that polyadenylation usage is biased in a tissue and disease related manner. Zhang and Tian (Zhang H et al., 2005) have also shown that the usage of polyadenylation sites as well as expression of protein factors involved in polyadenylation is likely to be different in brain tissues compared to other tissues (Zhang H et al., 2005). In spite of the evidence from large scale analysis of EST sequences, characterization of polyadenylation sites and 3’UTR variation in mammalian transcripts requires detailed experimental analysis. Experimental proof is lacking for majority of the predicted alternative polyadenylation events.

Here we used clues from hybridization patterns of probe sets in High Density Oligonucleotide (Affymetrix arrays) used for global profiling (Chapter-III) to identify potentially differentially expressed transcripts from the actin gene and confirmed it with experimental analysis and EST mapping. Further, analysis of microarray data and a survey of EST sequences showed that the cytoplasmic beta-actin gene in mouse may use two tandem polyadenylation sites that are used in a tissue specific manner. We carried out 3’RACE analysis to unambiguously establish the existence of two 3’UTRs of varying length in actin transcripts. The distal polyadenylation site is associated with a perfect PAS element (AAUAAA). Although a U-rich potential CstF binding site is present downstream to the proximal polyadenylation signal no
upstream canonical PAS element could be identified. On the other hand although the longer UTR containing transcript was expressed at a relatively lower level in neuronal cells, it conferred higher translational efficiency to the transcript and harbours miRNA target sites. Post-transcriptional regulation of expression from the longer UTR containing transcript is mediated, at least in part by a conserved mmu-miR-34b-5p/34a target site.

5.2 Results

The high density oligonucleotide array for mouse genome generated by Affymetrix, Mouse U74Av2 includes five probe sets complimentary to actin. Three probes (M12481_5_at, M12481_M_at and M12481_3_at) included in this and subsequent versions of the array are complimentary to 5’, middle and 3’ parts of the coding region. These probe sets are included on the array as control probes for housekeeping genes and are expected to show high level of signal after hybridization. Two additional probe sets (101578_f_at, 95705_s_at) are also annotated as corresponding to “cytoplasmic beta-actin gene (Actb)” but they are complimentary to regions 165 bp downstream to the reported 3’ end of the actin transcript (Fig 5.1). Each probe set consists of overlapping oligonucleotides that cover a part of the Actb gene. The sequences corresponding to the five probe sets are provided in Appendix (A).

In mouse neuronal gene expression profiling experiments carried out in our lab previously, we found that the two probe sets complimentary to the Actb downstream region also showed detectable signals. However the signal was much lower than the signals from the three control probes. This provided an indication that Actb gene may be associated with heterogenous 3’ UTRs. We examined the signals from all five probes in microarray data generated using U74Av2 arrays from Gene Expression Omnibus (GEO), the microarray data repository hosted by NCBI. There are more than 4000 experiments using the GPL81 platform corresponding to the U74Av2 in GEO. We observed that the expression of the three Actb internal probes were consistently higher than the signals from the two downstream probe sets. To explore tissue specific expression differences between the probes, we used the tissue specific expression data profiling carried out by Freilich S et al. (Freilich S et al., 2005).
Alternative polyadenylation in the cytoplasmic $\beta$-actin gene in mouse neuronal cells

Fig 5.1: Mouse cytoplasmic beta actin (Actb) specific antisense hybridization probe sets were mapped on the Actb sequence, a pictorial depiction. M12481_5_at, M12481_M_at, M12481_3_at, 95705_s_at, 101578_f_at are Actb specific antisense probe sets in Affymetrix U74Av2 mouse Genechip. M12481_5_at, M12481_M_at, M12481_3_at are complementary to the coding region of Actb transcript while 95705_s_at, 101578_f_at are complementary to the 3'UTR region. The nucleotide sequence corresponding to each probe sets are given in Appendix (A).
We found that the three internal probes give a consistently high expression level whereas the 3’UTR specific probes showed highly variable expression in different tissues (Fig 5.2). In tissues like pre-putial gland, pyloric andrus and vas deferens it showed a signal comparable to the high signal of the internal control probe sets, in tissues like appendix, brain and testes medium level expression was seen whereas extremely low signals were seen in colon and lung. This observation also ruled out the possibility that the 3’UTR specific probes simply hybridize less efficiently than the internal probes.

We next carried out EST analysis to validate the results from microarray analysis. ESTs reported from the actin gene locus were systematically analyzed to identify longer transcripts which may account for the signal from the downstream probes in the microarrays. The EST analysis revealed two distinct clusters of EST ends in the 3’UTR of Actb gene, corresponding to the two alternative polyadenylation sites (Fig 5.3A). The former cluster comprised of a total of 63 ESTs and the latter 122 ESTs. Members of each cluster were further analyzed for the tissue of origin using the Unigene Annotations (Fig 5.3B). Brain tissue showed enrichment for the first cluster of ESTs while the latter showed enrichment in the embryonic tissues and mammary gland.

We experimentally verified the presence of alternative 3’UTRs in cytoplasmic beta-actin transcripts. Total RNA isolated from Neuro-2a cells was reverse transcribed to synthesize cDNA and PCR amplification using a common forward primer and alternative 3’UTR specific reverse primers revealed the presence of transcripts of varying abundance (Fig 5.4A). The common forward primer spans the exon junction between the last and penultimate exons. One of the reverse primers is complimentary to the region 2 bp following the stop codon and results in a short 190 bp product. This primer can bind to cDNA products from both short and long transcripts. A second reverse primer that binds to the region 259 bp downstream to the stop codon was also included in the analysis. This primer is positioned beyond the predicted proximal polyA site and can bind only to the long transcripts.
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

Fig 5.2: Tissue specific expression of cytoplasmic beta actin (Actb). Microarray data (Freilich S et al., 2005) was reanalyzed and normalized values of Actb specific probes were plotted. Antisense probe sets corresponding to coding region (M12481_5_at, M12481_M_at, M12481_3_at) and 3’UTR (95705_s_at, 101578_f_at) of Actb are obtained from Affymetrix U74Av2 mouse Genechip.
Fig 5.3: EST end analysis of Actb transcript revealed two distinct clusters of transcripts corresponding to the two polyadenylation sites (A) The tissue of origin of ESTs in each cluster (B) showed that the ESTs were differentially expressed in different tissues. The scale bar represents abundance of ESTs.
Fig 5.4: Reverse transcription followed by PCR (RT PCR) (A) and 3’RACE (Rapid amplification of cDNA 3’ends) analyses revealed the existence of short (B) and long polyA variant (C) of Actb in mouse neuronal cells. Forward (FP) and reverse (RP) primers (arrow heads) for PCR amplification in RT PCR analysis were shown by pictorial depiction (A) where FP was an intron spanning primer common for both short and long transcript, E5 and E6: Exon 5 and 6 respectively, intronic region between E5 and E6 not shown. Gene specific FP and adapter primer for 3’RACE analysis were described in materials and method. The intensity of each band in A, left panel was quantified and plotted (A, right panel) relative to long UTR specific band. Data represents mean±SD of two independent experiments performed. (D) 3’RACE products were sequenced and 1st and 2nd cleavage sites (▼), FP used in RACE for amplifying long UTR (→), Affymetrix probe sets 95705_s_at (underlined) and 101578_f_at (dotted line), Zip code (italics) demarcated on the 3’UTR sequence. Nucleotide that differs from Genebank sequence was indicated as (/). M: DNA marker, S & L: transcript bearing short 3’UTR and long 3’UTR, respectively. RT+: with reverse transcriptase, RT-: without reverse transcriptase.
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

The 450 bp product arising from this primer reflects the abundance of the long transcript, whereas the abundance of the short 190 bp product represents the collective expression level of short and long transcripts from the Actb gene. These RT-PCR results, in agreement with the microarray and EST analysis showed, without doubt, that mouse neuronal cells express 3’UTR variants of Actb transcripts. To unambiguously assign the exact 3’ end of the Actb gene, we used the classical 3’ rapid amplification of cDNA ends (RACE) methodology. We carried out 3’ RACE for actin transcripts using total RNA isolated from the mouse neuronal cell line Neuro-2a. The total RNA was used for cDNA synthesis primed by an oligodT primer carrying a 5’ GC-rich adapter primer. The extended region of the adapter primer serves as a reverse primer binding site for subsequent rounds of PCR amplification using gene specific forward primers and a common reverse primer. Subsequently, the PCR products were cloned and sequenced. We isolated two distinct types of clones in 3’ RACE experiments that allowed us to unambiguously establish the existence of two alternative transcription stop sites in the Actb gene. The proximal stop site results in a short 3’ UTR of 59 bp that is highly conserved from chicken to human and includes the 54 nt zipcode sequence (Fig 5.5). A gene specific forward primer positioned 179 bp upstream from the translation stop site, resulted in the ~290 bp short product (Fig 5.4B) expected from the proximal polyadenylation site. However, in mouse neuronal cells we also found, additionally, actin transcripts with a much longer 3’UTR of 678 bp. The gene specific forward primer was designed to bind a region between the two polyadenylation sites so that it would amplify only the cDNA products of the long transcripts and result in a ~550 bp product (Fig 5.4C). The longer transcript includes regions homologous to the two downstream probes (101578_f_at, 95705_s_at) on the Affymetrix array. Figure 5.4D provides an overview of the sequence elements in the 3’UTR of the actin transcript along with the positions of the Affymetrix probe sets mentioned here, and the polyadenylation sites identified in our RACE experiments.
Fig 5.5: Conservation of Zip code sequence among different animals. 54 nt long Zip code sequence described in chicken beta actin (Kislauskis, E.H et al., 1994) has been aligned with beta actin sequence of different vertebrates using Multalin.
Our RT-PCR analysis implied that the longer transcripts were less abundant than the short transcript. However, to rule out the possibility that the PCR product abundance may be affected by the efficiency of PCR, we carried out in vivo expression analysis. We performed in situ hybridization using probes specific to the long UTR and common to both transcripts in neuronal cells to look for differential localization of the longer actin transcript. As shown in Fig 5.6, both long transcript specific antisense probe and the common antisense probe showed cytoplasmic localization while sense probes failed to show any signal. The signal from the common antisense probe that can bind to both long and short transcripts was consistently higher than the signal from the long transcript specific probe. Therefore, in keeping with the microarray data, the longer transcript seems to be less abundant in neuronal cells.

We cloned the short and the long UTR regions (Fig 5.7A) isolated in our RACE experiments downstream of the luciferase reporter gene driven by the highly and constitutively expressed CMV promoter. The luciferase activity per unit total protein showed that the long 3’UTR region conferred 2.4 fold higher steady state expression level of the reporter as compared to the vector transfected controls (Fig 5.7B). The expression level of the reporter from the short UTR-reporter fusion construct was reduced to nearly half of the vector control. The relative difference in luciferase activity between the short and long 3’UTR clones was 4.3 fold. The additional 515 bp in the longer transcript seems to harbour sequences that stabilize the transcript or improve translational efficiency, since protein levels from the reporter gene are significantly higher in the cells expressing the long transcript fused to the reporter gene.
Fig 5.6: Existence of polyA variants of Actb in differentiated mouse neuronal cell analyzed by in situ hybridization. Differentiated Neuro-2a cells were hybridized with digoxigenin-UTP labeled antisense RNA probe-a (A), sense probe-a (C), antisense RNA probe-b (B), sense probe-b (D). Antisense RNA probe-a: antisense probe that can detect both short and long polyA variants; Antisense RNA probe-b: antisense probe specific for long polyA variant. (scale bar, 100um).
Fig 5.7: Effect of short and long 3’UTR sequence on translation. (A) Sequence region of short and long 3’UTR of Actb was cloned (see materials & method) in a pmirReport™ vector (Ambion) containing luciferase as a reporter. Cloned region of short and long UTR were pictorially described (A) by primer sets where reverse primer was adapter primer (AP) used for RACE; short and long UTR specific forward primers were denoted by S and L, respectively, followed by arrow heads. (B) Translational effect of UTR sequences were determined by luciferase assay, 24H after transfection cells were lysed and luminescence were counted in a luminometer. Data was normalized to total protein and represented relative to vector transfected control. Data shown are mean ± SEM of four independent experiments, in replicate performed.
We next looked for potential microRNA target sites in the 3’UTR region of Actb. Using Miranda we predicted mmu-miR-34b-5p and mmu-miR-505 binding sites (Fig 5.11) in the long UTR of Actb transcript, downstream to the proximal polyadenylation sites. On closer inspection it was seen that the mmu-miR-34a and mmu-miR-34b-5p had identical seed sequences and therefore both could potentially target the same site. We checked the evolutionary conservation at these sites, since functional miRNA target sites in the 3’UTR are likely to show higher conservation than regions that do not harbour functional elements. The mmu-miR-34b-5p/34a target site was highly conserved amongst vertebrates (Fig 5.8A), while the mmu-miR505 site in the human homolog is deleted (Fig 5.8B). Due to poor conservation of the target site, mmu-miR-505 was not validated further. In summary, the longer transcript arising from the distal polyadenylation site seems to be less abundant, translationally more efficient and potentially targeted by mmu-miR-34b-5p/34a.

We attempted to study the role of miRNA mediated regulation on the long UTR of Actb by over expressing the pre-miRNA. We expressed pre-miR-34b under the CMV promoter in Neuro-2a cells along with the luciferase-UTR fused constructs. Cells expressing the miRNA and long UTR construct showed a modest up regulation (statistically significant pvalue < 0.05, Student’s t-test) of luciferase reporter. The long UTR in the antisense orientation was not affected (pvalue > 0.05, Student’s t-test) as compared to mock (pSilencer negative control vector) transfected control (Fig 5.9). There are several potential reasons for the lack of pronounced effect observed after transient expression of the pre-miRNA. Processing of pre-miRNA is known to be influenced by flanking sequences and inefficient processing of the pre-miRNA,due to sub-optimal flanking sequences can result in low induction levels. Both 5’ and 3’arms of the pre-miR-34b can give rise to mutually complementary miRNAs (mmu-miR-34a-5p and mmu-miR-34b-3p, respectively). However, the influence of the miRNA arising from the opposite arm has not been studied. Inefficient processing and or interference from mmu-miR-34-3p are some of the potential reasons for the limited effect on target gene. We also notice interference from the pSilencer vector that could not be explained at this stage. Next, we used anti-miRNA molecules to study the effect of down regulation of endogenous mmu-miR-34a/34b-5p. The high sequence similarity between miR-34a/34b-5p sequences allows the anti-miRNA molecule to target both mmu-miR-34a and mu-miR-34b-5p. Anti-miRNA molecules resulted in
up to 3 fold down regulation of expression from long UTR whereas short UTR remained unaffected (Fig 5.10). Endogenous mmu-miR-34a/34b-5p, therefore, mediates higher expression from the long UTR. Fig 5.11 provides a schematic representation of the sequence features identified in the actin 3’ UTR sequence, including localization signals, miRNA binding sites experimentally identified polyadenylation sites and associated sequence elements.
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

Fig 5.8: Conservation of mmu-miR-34b/a (A) and mmu-mir-505 (B) target regions among different vertebrates. Predicted (using Miranda) mmu-mir-34b-5p, mmu-miR-34a and mmu-mir-505 target regions on mouse cytoplasmic beta actin (Actb) sequence were aligned with human and rat Actb sequences using Multalin. Potential target region of mmu-miR-34b-5p and mmu-miR-34a are identical.
Fig 5.9: mmu-miR-34b dependent regulation of long 3’UTR variant of Actb in Neuro-2a cells. Long UTR specific region was cloned into pmirReport vector (Ambion) that contains luciferase as a reporter under CMV promoter. Long UTR was cloned in anti-sense orientation into the same vector. Pre-miRNA of mmu-miR-34b was cloned into pSilencer 4.1-CMV neo vector (Ambion). Neuro-2a cells were cotransfected with long UTR construct/ long UTR anti-sense orientation construct/empty vector along with pre-miR-34b construct/mock. pSilencer vector (Ambion) was used as negative control (mock). 24H after transfection cells were harvested for assays. Reporter (luciferase) activity was normalised to the activity of a cotransfected CMV-β-galactosidase construct. Data represented the mean ± SEM of two independent experiments, each in duplicate, performed.
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

A

LNA complementary

mmu-miR-34b-5p
mmu-miR-34a

B

Normalized Luciferase count

Vector Long UTR Short UTR

0 2000 4000 6000 8000 10000 12000 14000 16000

mock (40nM) LNA (40nM) mock (40nM) LNA (40nM) LNA (20nM) LNA (40nM) mock (40nM) LNA (40nM)
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

Fig 5.10: Inhibition of mmu-miR-34a/b by transfection of antisense locked nucleic acid (LNA) oligonucleotide leads to the down regulation of long 3'UTR variant of Actb in a concentration dependent manner (B). Antisense LNA oligonucleotide was designed against both mmu-miR-34b-5p/34a (A); and co-transfected with UTR constructs of Actb and empty vector into Neuro-2a cells. A non specific LNA was used as a negative control (mock). 24H after transfection cells were harvested for assays. Reporter (luciferase) activity was normalised to the activity of a cotransfected CMV-β-galactosidase construct. Data represented the mean±SEM of two independent experiments, each in duplicate, performed. (*) denotes modified base in LNA.
Fig 5.11: Annotation of 3'UTR of mouse cytoplasmic beta actin (Actb) reference sequence along with short and long polyadenylation variants. PAS: Polyadenylation signal, pA: Polyadenylation site.
5.3 Discussion

Recent bioinformatics studies have shown that nearly fifty percent of genes in the human genome may harbor alternative polyadenylation sites (Yan J and Marr TG, 2005). Heterogeneity in 3’UTRs can therefore account for differential regulation, stability and spatio-temporal expression of transcripts from the same gene. We have used EST and gene expression data, 3’ RACE analysis and in situ hybridization to establish that the cytoplasmic beta-actin gene generates two alternative transcripts terminated at tandem polyadenylation sites. Our analysis indicates that the longer transcripts are regulated in a tissue specific manner, whereas the proximal polyadenylation site is used for constitutive expression. Functionally, the longer UTR confers higher stability on the transcript, resulting in higher reporter gene expression, in mouse neuronal cells. Besides its highly conserved housekeeping function, the Actb gene plays diverse roles. It has been suggested that variable 3’untranslated regions may contribute to the complexity of transcriptome and provide a mechanism for differential regulation in a tissue specific manner (Zhang H et al., 2005). The existence of tandem polyadenylation site and heterogeneous 3’UTRs in the well studied Actb gene provides experimental support for the involvement of 3’UTR in tissue specific gene expression.

We explored the sequences surrounding the polyadenylation site for conserved motifs that have been shown to flank the cleavage site. Polyadenylation is carried out by the polyA polymerase in mammalian cells in a highly regulated manner, through a complicated mechanism involving multi-protein complexes that interact with these sequences on either side of the cleavage site. The most well characterized sequence element that signals polyadenylation sites is the AAUAAA motif found within 10-30 nt upstream of the cleavage site (Proudfoot, N., 1991; Colgan DF and Manley JL, 1997). The distal polyadenylation site is associated with a canonical PAS element followed by downstream G/U rich sequences and has been bioinformatically identified in whole genome studies. We did not find a canonical polyadenylation element upstream of the short transcript. The AAUAAA sequence can account for only about 53% of all the known polyadenylation sequences in the genome (Tian B et al., 2005). The major sequence variant of the PAS element, AUUAAA, accounts for an additional 16% of all polyadenylation sites predicted by EST mapping (Tian B et al., 2005). Several genes harbour other variants of the polyadenylation site, although
they vary in strength and maybe enhanced by the presence G/U or U rich sequences downstream. Nearly 13 most frequently occurring variants of the PAS site have been identified by large scale analysis of polyadenylation sites in the human and mouse genomes (Beaudoing E et al., 2000; Tian B et al., 2005). Although the canonical PAS element is not seen with 40nt upstream of the proximal polyadenylation site, we found a UUUACA hexamer at 31nt upstream of the proximal site. This hexamer has a single substitution at the 5th position from the hexamer UUUAAA which is amongst the ten most frequently found hexamers reported in previous studies. There is a strong U rich region 35 nt downstream of the proximal cleavage site that may enhance use of the non-canonical upstream signal.

EST and microarray analysis indicated that the long transcript is highly expressed in embryonic and germinal tissue, since detection levels were high in pre-putial gland, vas deferens and testes (microarray analysis) and embryonic tissue and mammary gland (EST analysis). The whole brain showed a medium level of expression of the long transcript in microarray and EST data whereas it was poorly expressed in other tissues. The inclusion of several cell types in whole brain samples and the resultant heterogeneous nature of whole brain RNA can result in dilution of expression level of transcripts that maybe expressed in a specific sub-population of neuronal cells. The actin transcript has been shown to undergo rapid deadenylation during early developmental stages in mouse embryos (Bachvarova R et al., 1985; Paynton BV et al., 1988). In neuronal cells the actin transcript is localized to the synapses as part of ribo-nucleoprotein bundles (Tiruchinapalli DM et al., 2003). Transient localization of actin transcripts following exposure to neutrophins and brain derived neurotrophic factor (BDNF) indicates that it has an important step in directed growth of axons and synaptic connectivity (Eom T et al., 2003). The localization of actin transcript requires a sequence element called the zipcode (Kislauskis EH et al., 1994). Reporter beta-galactosidase fusion to the zipcode results in localization of lacZ expression to sites of actin polymerization at the periphery of chick fibroblast cells (Kislauskis EH et al., 1994). The zipcode is a conserved 54 nucleotide long sequence that occurs in chick, dog, mouse and human actin transcripts, immediately following the stop codon. Both the short and long UTRs identified in our study, contain the zipcode sequence. The heterogeneity in the untranslated region apparently does not disrupt the localization signal. However, the structural basis of the binding of the Zipcode

121
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

Binding factor (ZBF) to the localization signal has not been studied. The additional regions in the longer transcript could potentially alter the secondary structure of the transcript and modulate the localization of actin mRNA to the cell periphery. Our in situ hybridization results serve to demonstrate the in vivo presence of the longer transcript in neuronal cells and provide a basis for studying specific localization of these transcripts to growth cones of axons following treatment with external agents like neutrophins and BDNF.

The longer untranslated region harbours target sites for regulation by small regulatory RNAs. MicroRNAs are 21-23 nt long regulatory RNAs that can bind to target sites of partial complimentarity within untranslated regions and interfere with transcript stability or translational efficiency. Alternative polyadenylation at tandem sites can result in transcripts coding for the same protein with differential susceptibility to miRNA mediated regulation at the translational level. We found that the longer actin transcript harbours potential binding sites for mmu-miR-34a/34b-5p and mmu-miR-505. The human, mouse and rat transcripts show high level of conservation at the miR-34a/34b-5p binding sites, implying that these sites are functional. We observed a pronounced dose dependent down regulation of luciferase reporter activity in cells transfected with LNA modified anti-miR-34a/34b-5p oligonucleotides and luciferase-long UTR fusion construct. Taken together, these results allow us to attribute a mmu-miR-34b-5p/34a mediated positive effect on expression from the long UTR of Actb.

miRNAs are generally believed to have a repressive effect on target molecules. Although rare, miRNAs have earlier been shown to mediate a positive regulatory effect in specialized conditions. For instance, the hsa-miR-369-3 binds to AU rich elements (ARE), in the 3’UTR of tumor necrosis factor-α (TNF-α) transcript, recruits specific proteins and activates translation under growth arrest conditions (Vasudevan S et al., 2007). A similar miRNA mediated up regulation of target transcript has been demonstrated by Let-7 at the 3’UTR of high mobility group A2 (HMG A2) (Vasudevan S et al., 2007). mmu-miR-34a/34b-5p mediated up regulation of Actb long UTR transcript is part of a growing body of evidence that miRNA can mediate repressive as well as inducing effect on gene expression.

In lower organisms like Drosophila the miR-34 family consists of a single miRNA whereas in human, it is represented by highly similar sequence variants hsa-miR-34a, b and c (Griffiths-Jones S et al., 2006). The human miRNA, hsa-miR-34a can be
induced by ionizing radiation and DNA damage agents in a p53 dependent manner (Bommer GT et al., 2007; Raver-Shapira N et al., 2007). It has been shown to be highly expressed in pancreatic cancers, actively participate in the p53 tumour suppressor pathway and target genes like Notch, Delta1 and the transcription factor E2F3, besides the anti-apoptotic gene bcl2 (Tarasov V et al., 2007; Hermeking H, 2007; Chang TC et al., 2007; Tazawa H et al., 2007). The existence of mmu-miR-34a/34b-5p target sites in the long UTR of Actb implies that further experimental analysis of differential expression of the 3’UTRs in cancers can be important. The actin transcript has been found in ribonucleoprotein complexes in neuronal synapses, but miR-34 was not found to be abundant in cell neurites, in the first study to identify miRNAs localized to the neurites (Kye MJ et al., 2007). Further studies of co-localization of the mRNA and miRNA in axons and dendrites of specific neuronal cells may reveal the functional relevance of the 3’UTR heterogeneity in Actb transcripts.

In summary our studies unambiguously establish the presence of tandem alternative polyadenylation sites and 3’UTRs of different tissue specific expression pattern in Actb mRNA transcripts. This is the first report of heterogeneity in transcripts of the Actb gene, widely held to be constitutively and ubiquitously expressed housekeeping gene. It supports the findings of several bioinformatics studies that uncovered the extensive occurrence of polyadenylation sites in the genome. Furthermore, this study illustrates that the integrated analysis of microarray data and EST data can provide novel insights into differential regulation and functional importance of alternative polyadenylation in the genome. The approach used here, of using differential hybridization signals from different probe sets to identify alternative splicing and polyadenylation can be used further for genome-wide bioinformatics analysis.

5.4 Materials and methods

5.4.1 EST and Microarray data analysis

Microarray data (accession id: E-HGMP-2, Freilich S et al. (Freilich S et al., 2005) was downloaded from Array express site. Data was normalized by using Z-score transformation method as described (Cheadle C et al., 2003).
Mouse ESTs were downloaded from the Unigene (IDs Mm.455830, Mm.391967, Mm.328431) and comprised of a total of 5195 ESTs derived from 43 tissues. 3’ read ESTs with Polyadenylation signal was parsed using Perl scripts as per the Unigene annotations. The final dataset of 3’read ESTs with polyadenylation signals comprised of 203 unique entries. The EST sequences were mapped back to the Genome location using BLAT using default options. The EST end positions and the tissues from which they were derived was parsed using Perl scripts with reference to the gene structure and positions (UCSC mm9).

5.4.2 RNA isolation

Total RNA was isolated from cultured Neuro-2a cells using TRIzol reagent (Invitrogen, USA) according to the manufacturer’s instruction. The RNA pellets were washed with 70% ethanol, centrifuged and dried. Pellets were re-suspended into DEPC treated water followed by the addition of 10X reaction buffer and RNase free DNase I (Fermentas). Samples were incubated at 37°C for 30 minutes. Then the RNA was cleaned using RNeasy Mini Kit (Qiagen) following the protocol by the manufacturer. RNA concentration and purity was determined by measuring optical density at 260 nm and 280 nm using a spectrophotometer (Eppendorf) and separating the RNA samples on a 1.5% agarose formaldehyde gel.

5.4.3 Rapid amplification of 3’-cDNA ends (3’ RACE)

cDNA was generated from total RNA sample by using adapter primer [AP: GGCCACGCGTCGACTAGTAC (T)_{17}] and M-MuLV reverse transcriptase (New England Biolabs Inc) at 42°C for 1 h. Subsequently RNaseH (Invitrogen) treatment was performed at 37°C for 20 min.

Touch down (TD) PCR was carried out using the following primers:

(1) GGCATTGCTGACAGGATGCAGAAGG (Gene specific primer)
(2) GGCCACGCGTCGACTAGTAC (Same as AP primer except T_{17} unit)

TD PCR condition was as followed: 94 °C for 3 min; 29 cycles of 94 °C for 1 min, 65 °C (with reduction per cycle 0.5°C) for 2 min, 72 °C (with elevation per cycle 0.2 °C) for 1 min; then 9 cycles of 94 °C for 1min, 55 °C for 2 min and 72 °C for 2 min; then final extension of 72 °C for 7 min.
Nested PCR was performed by using primer 2 (above) and the primer below:

GAAGGAGATTACTGCTCTGGCTCC [Nested primer for primer (1)]. Nested PCR condition: 95 °C for 5 min, 25 cycles of 94 °C for 30 sec., 60 °C for 30 sec., 72 °C for 2 min., then final extension at 72 °C for 10 min. In this way we were able to detect short 3’UTR.

Long 3’UTR was amplified by using touch down PCR product as template and Primer 2 (above) and long UTR specific primer mentioned below:

GTTTTGGCGCTTTTGACTCAGGATT

PCR condition was 95 °C for 5 min., 35 cycles of 95 °C for 10 sec., 60 °C for 30 sec., 72 °C for 30 sec., then final extension at 72 °C for 5 min.

PCR products were resolved on a EtBr stained 2% agarose gel. Purified PCR products were cloned in T-vector (pCR4®-TOPO®, Invitrogen) and confirmed by sequencing.

5.4.4 Constructs

Cloning of long and short UTR of Actb: The short and long 3’UTR of murine Actb was amplified from 3’RACE product synthesized from Neuro-2a cells. The following primers introduced Hind III (primer 1) and Sac I site (Primer 2 & 3) in the PCR products (restriction site underlined):

(1) Reverse primer: AAAAAGCTTGCCACGCGTCTAGTAC (common for Long and Short, contains sequence of adapter primer)
(2) Forward primer: AAAGAGCTCGTTTTGCGCTTTTGACTCAGGATT (for Long)
(3) Forward primer: AAAGAGCTCATCGTGCACCGCAAGTGCTTC (for Short)

The gel purified PCR products were cloned in Hind III and Sac I site of pMIR-REPORTTM vector (Ambion), in frame with Luciferase and confirmed by sequencing.

Cloning of long UTR in antisense orientation: The long UTR sequence was amplified from long UTR construct (mentioned above) using the following primers:

Forward primer: AAAAAGCTTGTTTTGCGCTTTTGACTCAGGATT (HindIII recognition sequence underlined)
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

Reverse primer: AAAGAGCTCGGCCACGCGGCGACTAGTAC (SacI recognition sequence underlined)

The gel purified PCR product was cloned in HindIII and SacI site of pMIR-REPORT™ vector (Ambion), in frame with Luciferase and confirmed by sequencing.

Cloning of pre-miRNA: Pre miRNA sequence of mmu-mir-34b (accession ID MI0000404) was taken from miRBase. The following overlapping primers were used for PCR amplification of pre-miRNA sequence of mmu-mir-34b:

Forward primer:
TAATCTAGAGGAATCCGTGCTCGGTTTGTAGGCAGTGAATTAGCTGATTGTAGTGCGGTGC
(BamHI recognition sequence underlined)

Reverse primer:
ATTGAGCTCAAAGCTTGCCTTGTGGTTTTGATGGCGAGTTAGTGATTGTCAGCAGCAGTACATCACATCTAACATTA
(HindIII recognition sequence underlined)

The gel purified PCR product was cloned in Hind III and BamHI site of pSilencerTM 4.1-CMV neo vector (Ambion) and confirmed by sequencing.

5.4.5 Cell culture, differentiation

Murine neuroblastoma cells (ATCC number CCL-131; Neuro-2a or N2a) were maintained in Minimum essential medium (MEM) (GIBCO-BRL) supplemented with 10% fetal calf serum [Biological Industries, Israel], 2mM L-glutamine [Sigma], 1mM sodium pyruvate [Sigma] and antibiotic-antimycotic solution (100X stock) [Sigma] at 37°C humidified incubator with 5% CO₂.

Differentiation: Approximately 8×10³ cells were seeded in each well of 2 chambered slide (Lab-Tek slide Nunc). Cells were maintained in Opti-MEM (GIBCO-BRL). After 24 hour media was replaced by OptiMEM with 5 mM N⁶,2’-o-Dibutyryladenosine-3’,5’-cyclic monophosphate sodium salt (dbcAMP) (Sigma) and incubated up to 4-5 days till differentiation.
5.4.6 *In situ* RNA localization

Probe preparation: Digoxigenin-UTP labeled RNA probes were prepared by in vitro transcription (IVT) using DIG RNA labeling Kit (T7) (Roche) following manufacturer protocol. Qiagen column purified PCR products were used as a template for IVT reaction. The following primers were used for PCR amplification of UTR regions that was used in IVT reaction for making sense and antisense probes: T7 promoter sequence has been underlined, in the primer sequences.

Primers used for detection of both short and long UTR:

**RNA probe-a**
Anti-sense:
FP: CCAACCGTGAAAAGATGACC
RP: TAATACGACTCACTATA\(\text{-}\)GGGTAACAGTCCGCTAGAAGCA
Sense:
FP: TAATACGACTCACTATA\(\text{-}\)GGGCCAACCGTGAAAAGATGACC
RP: TAACAGTCCGCTAGAAGCA

Primers used for detection of long UTR:
RNA probe-b
Anti-sense:
FP: GTTTTGCGCTTTTGACTCAGGATT
RP: TAATACGACTCACTATA\(\text{-}\)GGGACCAAAGCCCTCATACATCAAGTT
Sense:
FP: TAATACGACTCACTATA\(\text{-}\)GGGTTTTTGCGCTTTTGACTCAGGATT
RP: GACCAAAGCCCTCATACATCAAGTT

Differentiated cells were fixed with 4% paraformaldehyde for 10 min, washed 4 times for 5 min with phosphate buffer saline (PBS) supplemented with 0.1% Tween-20 (PBST). Cells were then incubated in pre-hybridization buffer (50% formamide, 5X SSC, 50ugml\(^{-1}\) yeast t-RNA, 50ugml\(^{-1}\) heparin, 0.1% Tween-20) for 1hr at 50°C. Digoxigenin- label RNA probe was then added and incubated overnight at 50 °C. Subsequently, following washes were given: 5 times with wash1 buffer (2X SSC, 50% formamide) each with 20 min; one wash with wash2 buffer (2X SSC, 25mgml\(^{-1}\) \(^{1}\)Rnase A) for 30 min at 37°C; followed by five washes (where SSC buffer was diluted in PBST in a stepwise manner) at room temperature for 15 min. Among those five
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

washes first wash (2X SSC, PBST) was given once, second wash (1X SSC, PBST) and third wash (0.1X SSC, PBST) were performed twice. Cells were kept in blocking solution (2 mg/ml BSA, PBST) for 1 hr at room temperature and subsequently, incubated with alkaline phosphatase conjugated anti-digoxigenin antibody (Roche) (1:300 dilution in PBST) for 30 min at room temperature. After that twice wash was given with wash buffer (100mM Tris-Cl, pH 7.5, 150mM NaCl) for 15 min at room temperature. Hybridized probes were then detected with NBT/BCIP solution. After a 5 min wash with PBS cells were observed under 60X objective (bright field) in an inverted microscope (NIKON).

5.4.7 Transfection, luciferase assay

Approximately $3 \times 10^5$ cells (Neuro 2a) were seeded in each well of 6 well plate (Axygen) 24 h prior to transient transfection when 50-80% confluency was reached. Cells were washed once with Opti-MEM (GIBCO-BRL) and maintained in 1 ml Opti-MEM. Equivalent amount of transfection quality plasmids of vector alone (pMIR-REPORT™), cloned long UTR and short UTR of Actb (Endo free plasmid maxi kit; Qiagen) were used for transfection using Fugene 6 reagent (Roche) following manufacturer’s protocol.

24 hours after transfection cells were lysed in CCLR buffer (Promega) and total protein was estimated using BCA protein estimation kit (Sigma). Luciferase assay was performed using luciferase assay system (Promega) and luminescence was measured in a microplate luminometer (Berthold detection system). For data analysis luminescence values were normalized to total protein.

5.4.8 miRNA target prediction and validation experiments

miRNA target prediction was carried out using the Miranda software (Betel D et al., 2008 36(Database issue):D149-D153). We used the long UTR sequence as the input and a cut-off of predicted $\Delta G < -20$ kcal/mol.

Neuro-2a cells were co-transfected with luciferase-long UTR fusion construct and cloned pre-miR-34b. β-gal plasmid was also co transfected for normalization control. Luciferase-long UTR in antisense orientation fusion construct/empty vector were used as negative control. Transfection was performed using Amaxa nucleofection method as per manufacturer’s protocol.
Alternative polyadenylation in the cytoplasmic β-actin gene in mouse neuronal cells

For LNA experiments approximately $10^5$ Neuro-2a cells per well of a 12 well plate was used for transfection. Neuro-2a cells were co-transfected with luciferase-long UTR fusion construct/correspondingly short UTR construct/empty vector along with antisense LNA oligonucleotide (10nM-40nM) /mock LNA oligo (40nM) and β-gal control plasmid using NeoFX™ (Ambion) transfection reagent following manufacturer’s protocol. Due to the high level of sequence similarity the antisense LNA could be able to target both mmu-miR-34b-5p and mmu-miR-34a. The antisense LNA sequences (modified base is underlined) are given below:

(1) ACAAACCAGCTAAGA
(2) ACCAATTCGACCAC (Mock LNA or non specific LNA)

24H after transfection cells were lysed in reporter lysis buffer (RLB) (Promega) and reporter assays were performed.

5.4.9 Reverse transcription followed by PCR (RT-PCR)

Total RNA was isolated from Neuro-2a cells transfected with luciferase-long UTR fusion construct/ luciferase-short UTR construct/empty vector that contains luciferase reporter. RNA isolation method and DNase treatment has been described above. Total RNA was reverse transcribed by using random hexamer and M-MuLV reverse transcriptase (New England Biolabs Inc). cDNA was then PCR amplified by using luciferase specific primers (FP: CTCGGGTGTAATCAGAAT, RP: TTTACACGAAATTGCTTCT) and 18S rRNA specific primers (FP: CTTTCGAGGCCCTGTAATTG, RP: CCTCCAATGGGATCCTCGTTA).