Functional implications of the emergence of alternative splicing in hnRNP A/B transcripts

SIEW PING HAN,[1] KARIN S. KASSAHN,[2,3] ADAM SKARSHEWSKI,[1] MARK A. RAGAN,[2] JOSEPH A. ROTHNAGEL,[1] and ROSS SMITH[1]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD 4072, Australia
[2]ARC Centre of Excellence in Bioinformatics and Institute for Molecular Bioscience, The University of Queensland, St. Lucia, QLD 4072, Australia

ABSTRACT

The heterogeneous nuclear ribonucleoproteins (hnRNPs) A/B are a family of RNA-binding proteins that participate in various aspects of nucleic acid metabolism, including mRNA trafficking, telomere maintenance, and splicing. They are both regulators and targets of alternative splicing, and the patterns of alternative splicing of their transcripts have diverged between paralogs and between orthologs in different species. Surprisingly, the extent of this splicing variation and its implications for post-transcriptional regulation have remained largely unexplored. Here, we conducted a detailed analysis of hnRNP A/B sequences and expression patterns across six vertebrates. Alternative exons emerged via the introduction of new splice sites, changes in the strengths of existing splice sites, and the accumulation of auxiliary splicing regulatory motifs. Observed isoform expression patterns could be attributed to the frequency and strength of *cis*-elements. We found a trend toward increased splicing variation in mammals and identified novel alternatively spliced isoforms in human and chicken. Pulldown and translational assays demonstrated that the inclusion of alternative exons altered the affinity of hnRNP A/B proteins for their cognate nucleic acids and modified protein expression levels. As the hnRNPs A/B regulate several key steps in mRNA processing, the involvement of diverse hnRNP isoforms in multiple cellular contexts and species implies concomitant differences in the transcriptional output of these systems. We conclude that the emergence of alternative splicing in the hnRNPs A/B has contributed to the diversification of their roles in the regulation of alternative splicing and has thus added an unexpected layer of regulatory complexity to transcription in vertebrates.

Keywords: hnRNP A/B; alternative splicing; evolution

INTRODUCTION

Splicing is the process by which pre-mRNA is converted to mature mRNA through the excision of introns and ligation of exons. This involves the recognition of *cis*-acting sequences, including 5′ and 3′ splice sites (ss) and other exonic splicing regulators (ESRs) or intronic splicing regulators (ISRs), by *trans*-acting splicing factors that regulate ss selection. Alternative splicing occurs when more than one mRNA product is formed from the same pre-mRNA and can arise through several mechanisms (Kim et al. 2008), including exonization of introns (Sorek 2007), the emergence of alternative 5′ss and 3′ss via the weakening of 5′ss, and changes in the ESR frequency of ancestral constitutive exons (Koren et al. 2007; Lev-Maor et al. 2007). Therefore, whether an exon is included in the mature mRNA is determined by the frequency and strength of multiple *cis*-elements that recruit *trans* factors to regulate splicing (Hertel 2008; Wang and Burge 2008).

As splicing repressors, the heterogeneous nuclear ribonucleoproteins (hnRNPs) A/B are important *trans*-acting factors in alternative splicing. They are also involved in multiple stages of RNA metabolism and play critical roles in RNA processing, mRNA trafficking, and telomere maintenance (Krecic and Swanson 1999; Dreyfuss et al. 2002; Smith 2004; He and Smith 2009). The family includes the paralogs A1, A2/B1, and A3, as well as A0, a very minor paralog encoded by an intronless gene, postulated to have originated from a retrotransposition event, which will not be included in this study (Myer and Steitz

[3]**Present address:** Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, St. Lucia, QLD 4072, Australia.
Reprint requests to: Ross Smith, School of Chemistry and Molecular Biosciences, The University of Queensland, Room 425, Building 76, Cooper Road, St. Lucia QLD 4072, Australia; e-mail: ross.s@uq.edu.au; fax: 61-7-33654699.
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2142810.

1995; Akindahunsi et al. 2005). The hnRNP A/B paralogs are structurally similar (Fig. 1A), but differ markedly in their alternative splicing patterns (Fig. 1B). To date, the terms A1 and A2 have been used to represent either combined or individual isoforms, depending on whether their alternative splicing has been taken into consideration. To prevent confusion, we will use A1* when referring to both A1 and A1b proteins, while A2 will strictly refer to the A2/B1 isoform that excludes exon *2*. The levels of exon inclusion for the homologous cassette exons *7b*, *9*, and *8* of A1*, A2/B1, and A3 vary between paralogs. While exon *2* of A2/B1 is an alternative cassette exon, exon *1b* of A3 arises from an alternative 5′ss; these exons do not have equivalents in A1* (Buvoli et al. 1990; Hatfield et al. 2002; Ma et al. 2002). Importantly, A1 modulates splicing of its own pre-mRNA by recognizing consensus sequences in the introns flanking its alternative exon (Chabot et al. 1997; Hutchison et al. 2002). Thus, the hnRNPs A/B are both regulators and targets of alternative splicing.

While the hnRNP A/B isoforms have been described, their functional and regulatory significance and evolutionary history have remained largely unexplored. In particular, species-specific variations in splicing patterns and isoform expression have not been investigated to date. In vitro assays have shown that the distal splice site selection, RNA annealing, and RNA binding properties of A1 differ from those of A1b (Mayeda et al. 1994). A2/B1 isoforms have different intracellular localization patterns, and A2b appears to be the predominant isoform involved in mRNA trafficking in neuronal cells (Han et al. 2010). In addition, the levels of B1 mRNA and protein are up-regulated in nonsmall cell lung cancer, while levels of the other A2/B1 isoforms remain stable (Wu et al. 2003; Sueoka et al. 2005). Finally, expression levels of A1* and A2/B1 isoforms vary between tissue types and across developmental and cell cycle stages (Kamma et al. 1999, 2001; Matsui et al. 2000; Hatfield et al. 2002; Maggipinto et al. 2004; He et al. 2005). These data suggest that the observed differences in alternative splicing are the result of distinct selection pressures acting on hnRNP paralogs and isoforms in various cellular contexts. Thus, further investigation of isoform-specific variation in the expression and function of the hnRNPs A/B is crucial for understanding of their roles in regulating multiple steps of mRNA metabolism.

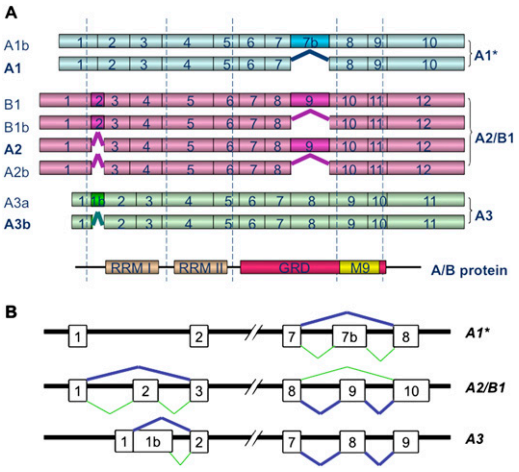


FIGURE 1. Structure of hnRNPs A/B at gene, transcript, and protein level. (*A*) Comparison of exons in A1*, A2/B1, and A3 in mouse mRNA and approximate alignment with protein domains. A1* is alternatively spliced to produce the transcripts *A1* and *A1b*, A2/B1 is spliced to produce transcripts *B1*, *A2*, *A2b*, and *B1b*, and A3 is spliced to produce transcripts *A3a* and *A3b*. Each numbered rectangle represents an exon, with lengths adjusted to align homologous exons. Vertical dashed lines highlight the structural similarity in terms of exon lengths, the correspondence of exons to functional protein domains, and the location of alternative exons relative to the entire mRNA. Major transcripts (A1, A2, and A3b) are in boldface. Alternative exons are in a darker shade. RRM, RNA recognition motif; GRD, glycine-rich domain; M9, nuclear localization domain. (*B*) Splicing of alternative exons in hnRNPs A/B in mouse. Boxes represent exons and black horizontal lines represent introns. Bold blue and faint green lines represent major and minor splicing patterns, respectively. Note that exon *7b* is mostly excluded from A1* exon *7b*, while the corresponding exon *9* in A2/B1 is generally included and exon *8* in A3 appears to be constitutively present in mouse.

In this study, we used computational and experimental analyses to reconstruct the evolutionary mechanisms that led to the variation in splicing patterns of hnRNPs A/B in six vertebrates. We show interspecies differences in splicing patterns and demonstrate the effects of alternative exons on protein function and expression. Our results indicate that the hnRNP A/B family of proteins has increased its functional diversity through the evolution of dynamic and highly regulated, species-specific alternative splicing patterns.

RESULTS

Identification of splicing regulatory *cis*-elements

The hnRNPs A/B are consistently more conserved across orthologs than across paralogs, suggesting that there has been limited concerted evolution within each species (Fig. 2). These results were further confirmed by phylogenetic analyses (Supplemental Fig. S1). Comparison of mouse *A2/B1* and *A3* showed that there is no similarity between their N-terminal alternative exons (AEs), indicating that these exons arose independently. These AEs are almost perfectly conserved within mammals (Table 1), suggesting that they may be functionally significant, while their absence from chicken and *Xenopus* sequences indicates that they arose only after the divergence of mammals.

Multiple alignments of the intronic regions flanking the C-terminal AEs revealed the presence of sequences corresponding to or resembling known A1* binding sequences (Chabot et al. 1997; Hutchison et al. 2002; Wang and Rothnagel 2004; Venables et al. 2008; Supplemental Fig. S3A). However, motifs resembling A1* binding sites were present only in rodent A2/B1 sequences, and in human A3

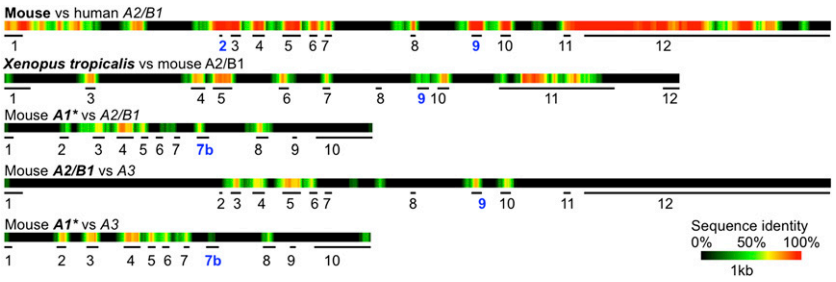


FIGURE 2. Pairwise alignment of hnRNP A/B genes in different species. Red and green regions correspond to sections of high and low identity, respectively. The black bars *under* the heatmaps represent the locations of exons within the genes. The exon numbering is based on that used in Figure 1, and the exon size and numbering correspond to that of the species or paralog in boldface. The numbers for alternative exons are in blue. Additional heatmaps can be viewed in Supplemental Figure S2.

sequences. Another key *cis*-element for splicing is the polypyrimidine tract, which is present in intron 1 of *A2/B1*, and is highly conserved between mammals. No such motifs are present in the corresponding introns in chicken or *Xenopus A2/B1* (data not shown).

Closer inspection of the sequences of the first exon of A3 revealed that exon *1/1b* in mammals has substantially expanded in size from the presumably ancestral sequence found in *Xenopus* (Supplemental Fig. S3B). In addition, the ss sequences of C-terminal AEs in the hnRNPs A/B vary between species, leading to differences in predicted ss strength (Fig. 3). For *A1**, there has been a major decrease in 5′ss strength in chicken and mammals compared with that in *Xenopus*, which is due to a single transversion from A to G at the +4 position (Fig. 3). The ss strength of *A2/B1* has also decreased, albeit to a smaller extent, whereas ss strengths and sequences vary widely between species for *A3*. Thus, splicing regulatory elements have diverged between species, which may have led to changes in alternative splicing patterns.

Quantification of alternative splicing patterns

To confirm the interspecies differences in splicing patterns suggested by our sequence analyses, we performed RT-PCR on brain cDNA from six species with primers that amplified across AEs (Fig. 4A). Two bands representing novel alternatively spliced isoforms in human and chicken (Fig. 4A, arrows) were excised and sequenced. We found, for the first time, that exon *8* of human *A3* is alternatively spliced at very low levels, and that chicken *A3* is alternatively spliced. Chicken *A3* has 3 exons upstream of the exon corresponding to exon *2* in the other species. The lower band on the gel corresponds to the amplicon that excludes the third of these exons (Fig. 4B).

In both chicken and *Xenopus*, N-terminal AEs were not present and C-terminal AEs were constitutively spliced for all paralogs (Fig. 4A). However, chicken *A3* was alternatively spliced between exons *1* and *2* in a pattern different from that of the mammalian species. While *A3* is alternatively spliced at exon *1b* for all mammals, exon inclusion is favored in human and cow, whereas exon exclusion is favored in rodents.

While exon *2* of *A2/B1* is alternatively spliced in all mammals studied, the ratio of amplicons including exon *2* was higher in cow than in the other mammals (Fig. 4A). Similarly, exon *7b* of *A1** was completely excluded in cow, which was unique within the mammals (Fig. 4A). Notably, the *CE1a* sequence of cow is the least conserved among the mammals, with a 6 base-pair (bp) deletion at the 5′ end (Supplemental Fig. S3A). In addition, the C-terminal AEs of *A2/B1* and *A3* are alternatively spliced only in rodents and human, respectively. These splicing patterns are consistent with the presence of A1* consensus sequences in the flanking introns (Supplemental Fig. S3). Thus, A1* may be important in the regulation of splicing of these alternative exons.

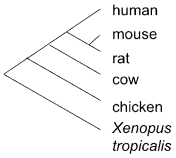
Effects of alternative exons on protein–mRNA interactions and protein expression

We next investigated the functional significance of alternative exons by comparing the affinity of hnRNP A/B isoforms for nucleic acid sequences (Fig. 5A). A2RE11 is a *cis*-acting element in certain mRNAs that directs their trafficking into neuronal dendrites, while Telo4 consists of four telomeric repeats. Both sequences bind the hnRNPs A/B with high affinity (Hoek et al. 1998; Shan et al. 2003; Moran-Jones et al. 2005; Yeo et al. 2005; Gao et al. 2008). In rat whole-brain lysate (WBL), A1, A2 and A3a are the predominant isoforms (Fig. 5A, left panel). In affinity pulldowns (PDs) using the A2RE11 and Telo4 sequences, A1 and A2 are still the predominant isoforms, but A3b is enriched to a greater extent than A3a (Fig. 5A, right panel). These results show that the inclusion of exon *1b* in A3 protein leads to a decreased affinity for its cognate mRNA sequences.

TABLE 1. Percentage conservation level of alternative exons in hnRNPs A/B

Alternative exon	A1	A2/B1	A3
N-terminal AE (across mammals)	NA	100% (36/36)	97% (64/66)
C-terminal AE (across mammals)	97% (155/159)	99% (119/120)	86% (126/147)
C-terminal AE (across all species)	42% (89/213)	77% (92/120)	62% (101/162)

Numbers in parentheses refer to number of conserved nucleotides out of total nucleotide positions; AE, alternative exon; NA, not applicable.



Exon 7b (*A1**)		Exon 9 (*A2/B1*)		Exon 8 (*A3*)	
3′ss	5′ss	3′ss	5′ss	3′ss	5′ss
90.8	83.59	86.98	84.71	92.96	88.44
90.8	83.59	86.98	84.71	97.45	103.24
90.8	83.59	86.98	84.71	Sequence unavailable	
90.8	83.59	86.98	84.71	93.16	68.71
Sequence unavailable		83.59	84.21	82.58	88.68
Uncertain	94.26	89.76	90.29	99.22	83.68

A1 exon 7b 5′ss sequences:

Mammal/chicken GTG | GTAGGT
Xenopus GTG | GTAAGT

FIGURE 3. Changes in splice site strength Splice site strengths of C-terminal AEs in different species. The 5′ss sequence flanking A1 exon *7b* differs between *Xenopus* and the other species at the +4 position (arrow).

The A3b transcript is the minor alternatively spliced isoform in humans, and no A3b protein was detected in human cell line lysate (Fig. 5B). In contrast, the A3a transcript is the minor alternatively spliced isoform in rodents (Fig. 4), but A3a protein was expressed at higher levels than A3b protein (Fig. 5A, left panel). This discrepancy between transcript and isoform ratios led us to examine whether the exon *1b* sequence could increase translational efficiency. Insertion of exon *1b* into A3ex1b-mGFP decreased protein expression levels by 15% compared with those of pEGFP-N1 (Fig. 5C), which was surprising since the presence of exon *1b* in rodent A3a was associated with enhanced protein expression. We inserted exon *1b* into pEGFP-2mer and pEGFP-4mer, two systems with lower translational efficiency, corresponding to ~25% and 10% of those of pEGFP-N1 (Wang and Rothnagel 2004). Inclusion of exon *1b* increased the expression levels of the resulting constructs, A3ex1b-2mer-mGFP and A3ex1b-4mer-mGFP, by ~70% (Fig. 5C). Next, we inserted exon *2* of A3 into pEGFP-2mer and pEGFP-4mer to test if the increase in translation was sequence specific. Exon *2* inclusion also increased expression levels, but to a lesser degree than inclusion of exon *1b*. Thus, it appears that the exon *1b* sequence of A3 represses expression from vectors with high translational efficiency, but enhances expression from vectors with low translational efficiency. This suggests that the emergence of alternative exon *1b* in mammalian A3 could affect regulation of A3 protein expression.

DISCUSSION

In this study, we have explored the evolution and significance of splicing variation in the hnRNPs A/B subfamily of proteins. We have identified novel splice transcripts and demonstrated species-specific variation in splicing patterns. Importantly, inclusion of different exons had significant

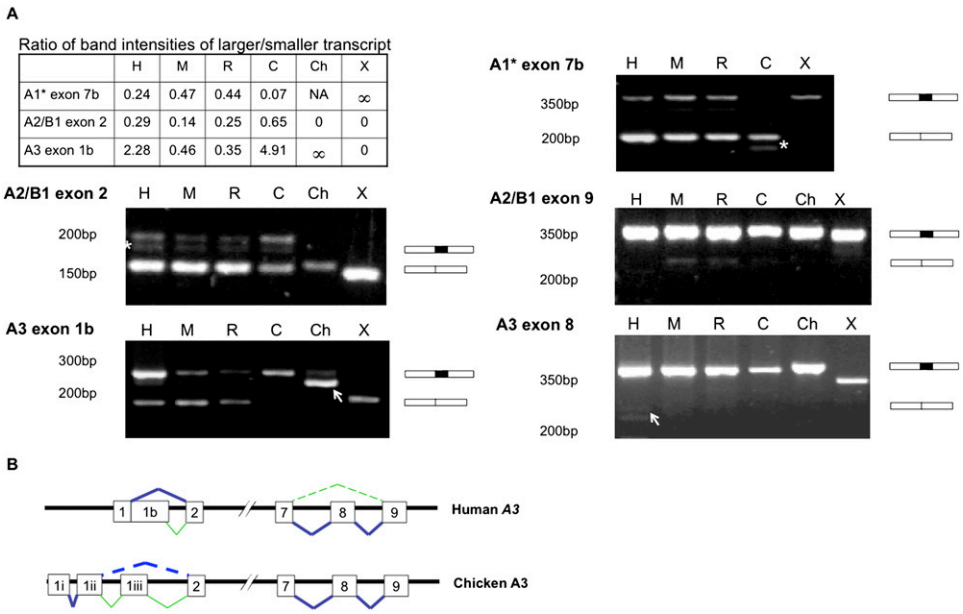


FIGURE 4. Expression patterns of alternative exons in different species. (*A*) RT-PCR analysis of inclusion levels of alternative exons of hnRNPs A/B in brain cDNA from different species (labels as in Fig. 3). Arrows point to novel amplicons not previously described, while asterisks indicate presence of double bands due to band-splitting. Ratios of band intensities of amplicons that have variable proportions across mammals are summarized in Table 1. (*B*) Splicing of alternative exons in *hnRNP A3* in human and chicken. Boxes represent exons and black horizontal lines represent introns. Bold blue and faint green lines represent major and minor splicing patterns, respectively. Novel splicing patterns identified in the RT-PCR analysis are dashed.

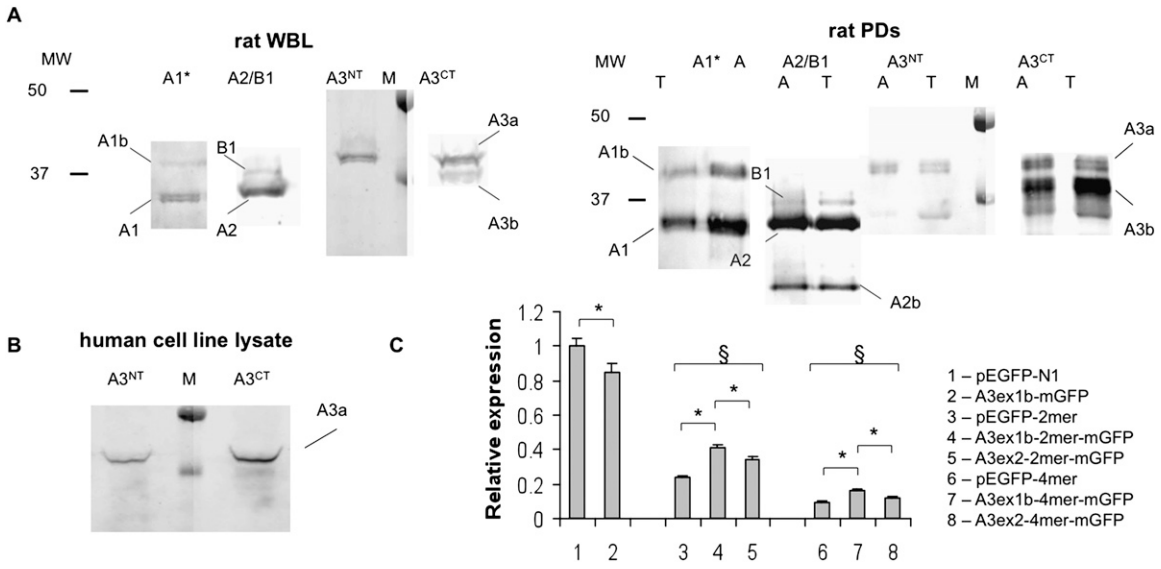


FIGURE 5. Effect of alternative exons on protein function and expression. (*A*) Binding properties: Western blots of rat whole-brain lysate (WBL), and A2RE11- and Telo4-binding proteins (A and T, respectively) isolated from rat brain by affinity pulldowns (PDs). The ratio of band intensities within a lane reflect the ratio of isoform protein levels present in the sample loaded. The samples for A3a and A3b were run on the same gel separated by a lane for the protein size marker (*M*), which has been included in the images to ensure accurate alignment of the lanes and correct assignment of isoform identities. Anti-A3^{NT} recognizes A3a only, while anti-A3^{CT} recognizes both A3a and A3b. (*B*) A3b is not expressed in human: Whole-cell lysate from SH-SY5Y cells, a human neuroblastoma cell line immunostained with anti-A3^{NT} and anti-A3^{CT}. A3b was not detected under these loading and staining conditions. (*C*) Translation efficiency: FACS measurements of pEGFP levels in HeLa cells transfected with pEGFP expression vectors containing exon *1b* or *2* of A3. §$P < 0.001$ by single factor ANOVA, *$P < 0.05$ by least significant difference method ($n = 3$). Bars represent standard deviation.

effects on nucleic acid binding affinity and translational efficiency. As hnRNPs are important regulators of RNA metabolism themselves, the expression of different hnRNP isoforms in varying cellular contexts implies associated changes in the transcriptional output of such cells. Thus, our findings highlight a new level of isoform- and species-specific variation in these key RNA processing proteins in vertebrates.

Evolution of alternative exons in hnRNPs A/B

We have shown that a variety of mechanisms have contributed to the evolution of alternative splicing in the hnRNPs A/B, including exonization (exon *2* of A2/B1), introduction of a new 5′ss (exon *1b* of A3), and a transition from constitutive to alternative exon inclusion (C-terminal AEs). These mechanisms appear to be mammal specific and would represent a means of increasing protein diversity without increasing gene numbers.

The variation in splicing patterns between paralogs and species are consistent with the presence and type of *cis*-regulatory elements, including ss, ISRs, and polypyrimidine tracts, and the interspecies variation in C-terminal AE inclusion levels could be attributed to differences in the frequency of A1* binding sequences and ss strengths. In particular, the low inclusion level of exon *7b* in A1* is uncharacteristic of alternative exons that evolved from a constitutive ancestor (Lev-Maor et al. 2007). The combination of a single base transversion that greatly weakened the 5′ss (Roca et al. 2005) and the accumulation of multiple A1* binding sequences likely resulted in the unexpectedly low inclusion rates of this exon.

There are currently two models for the evolution of alternative splicing. The first describes the emergence of new splice sites due to mutations in DNA sequences (*cis*-based), and the second the emergence of new splicing factors (*trans*-based) (Ast 2004). Both models apply to A1* evolution, wherein mutations gave rise to intronic A1* binding sequences that increased expression of A1, and the dependence of A1 expression upon binding of A1* to A1* binding sequences in turn selected for an increased frequency of these motifs. Thus, the ability of A1* to splice its own pre-mRNA may have accelerated the evolution of A1 toward higher expressed forms.

Variation in alternative splicing across species

Approximately 90% of minor-form alternative exons (inclusion level <1/3) were created during vertebrate evolution, and the rate of exon creation was particularly high during mammalian evolution (Alekseyenko et al. 2007). Here, we have shown that a similar pattern applies to mammalian hnRNPs A/B which have unique, minor-form AEs not found in *Xenopus* or chicken. In addition, there has been

further divergence within mammals, with splicing of exon *8* of A3 and exon *9* of A2/B1 occurring only in human and rodents, respectively, as shown in Figure 4. More than 11% of alternative exons are differentially spliced in human and mouse (Pan et al. 2005; Yeo et al. 2005), and 72% of minor-form exons are not conserved between their genomes (Modrek and Lee 2003). These species-specific differences are important, not only for understanding evolutionary and functional diversity, but also because they affect the extent to which experimental results obtained in one species can be extrapolated to other organisms with different or unknown splicing patterns. This is especially significant if different isoforms have different functional characteristics, as appears to be the case for the hnRNPs A/B.

Functional significance of alternative exons

The hnRNPs A/B are modular proteins that contain both RNA-binding and glycine-rich domains (Dreyfuss et al. 1993). The N-terminal AEs code for regions slightly upstream of the RNA-binding domains, while the C-terminal AEs code for regions within the glycine-rich domains (Hatfield et al. 2002; Ma et al. 2002). As these domains participate in or modulate interactions with nucleic acids (Mayeda et al. 1994; Maris et al. 2005), modification of their structural properties by the presence of alternative exons can generate isoform-specific differences in nucleic acid-binding properties, which in turn would impact upon multiple steps in nucleic acid processing. There is increasing evidence that the presence or absence of alternate exons in the hnRNP A/Bs modifies their function. First, insertion of exon *7b* into A1b decreases its splicing activity relative to A1 (Mayeda et al. 1994). Second, inclusion of exon *1b* into A3a decreases its affinity for the A2RE11 sequence (Ainger et al. 1997), which directs mRNA trafficking in neuronal cells (Munro et al. 1999). In addition, the A2/B1 isoform that excludes both exons *2* and *9* is the major cytoplasmic isoform in neuronal cells (Han et al. 2010), and hence should participate in mRNA trafficking to a greater extent than the other isoforms. Third, the presence of exon *1b* in A3a decreases its affinity for telomere repeats, which presumably has implications for its role in telomere maintenance (Tanaka et al. 2007).

In addition, we have shown that inclusion of exon *1b* of *A3* in pEGFP vectors increases pEGFP expression to a significantly greater extent than a sequence of identical length from a downstream exon. Exon *1b* of *A3* has an extremely high GC content of 80% and is located only 3 nucleotides (nt) downstream from the start codon, so it is well positioned to influence translation initiation. GC-rich regions generally reduce translational efficiency, but their effects may depend on the context and optimality of the start codon (Kozak 1991). Thus, exon *1b* of A3 may repress translation of open reading frames with an optimal context while enhancing those with suboptimal start codon context. However, exon *1b* increases expression levels by only 20%–30% relative to a downstream sequence of identical length, which indicates that other mechanisms, such as mRNA stability, are also contributing to the regulation of A3a expression.

Our work has revealed previously unrecognized interspecies variation in the splicing of hnRNPs A/B, as well as isoform-specific expression and function. As hnRNPs A/B participate in many aspects of mRNA metabolism, their functional diversification through the generation of alternatively spliced isoforms has potentially far-reaching implications for the entire transcriptome. Furthermore, the functional diversification of the hnRNPs A/B is particularly pronounced in mammals, which is likely to have contributed to the generation of increased post-transcriptional complexity. In conclusion, the hnRNPs A/B contribute an additional layer of complexity to species-specific regulatory control and the networks that underlie alternative splicing.

MATERIALS AND METHODS

Sequence source and alignments

Gene and mRNA sequences were downloaded from Ensemblv51 (Hubbard et al. 2009), Xenbase 2.3.1 (Bowes et al. 2010), and Entrez Gene (Maglott et al. 2005). Details of the sequence IDs and status are provided in Supplemental Table S1. To perform pairwise sequence comparisons between paralogs and species, we developed an alignment tool in which the degree of sequence identity is represented on a color scale. The resulting heatmaps allowed us to compare easily the level of conservation across entire genes and pick out regions of sequence conservation that may represent regulatory elements. Pairwise sequence alignment was performed using the Needleman–Wunsch algorithm (Needleman and Wunsch 1970) using a gap-opening penalty of −50, no-gap extension penalty, +5 for a nucleotide match, and −4 for a mismatch. Vectors containing percentage identity values were generated from each pairwise alignment. Percentage identity values for each base of the genes were calculated using the following equation:

$$\%\text{Identity} = \frac{\text{Number of matching bases in window}}{\text{Size of window}},$$

with the window defined as all bases (and gaps) in the pairwise alignment that are less than half the maximum window size from the base of interest. Vectors generated using a maximum window size of 75 were converted to heatmap diagrams by correlating percentage identity with color (heat) gradients. Multiple sequence alignments were performed using Clustal W2 (Larkin et al. 2007) and the default parameters.

Splice site scores and motif frequencies

The strengths of 3′ss and 5′ss were compared using the online Analyzer Splice Tool. This tool calculates the strength of 5′ss and

3′ss based on their identity to the consensus sequences CAG|gtaagt and ttttttttttcag|G, respectively (Shapiro and Senapathy 1987). For uniformity, the settings used for rodents were used for calculations in all species. Splice site strengths were also calculated using MaxEntScan (Yeo and Burge 2004) employing four different models. The various methods produced congruent results.

The ESRsearch tool, available at http://ast.bioinfo.tau.ac.il/ESR.htm, was used to search for putative exonic splicing enhancers (PESEs) and putative exonic splicing silencers (PESSs) described by Fairbrother et al. (2002). These are hexamers conserved at the wobble position between human and mouse sequences, which occur at a higher than expected frequency based on alternative codon usage and are thought to recruit splicing regulatory proteins. PESEs and PESSs were considered to be part of an exon if four or more bases were within the exon. As PESEs/PESSs are short and degenerate, there were many cases in which several PESEs/PESSs overlapped the same stretch of sequence. For instance, the sequence shown in Scheme 1

contains five PESEs/PESSs (underlined), although their potential to recruit regulatory proteins is unlikely to be five times that of a single PESE/PESS. To avoid overestimation of the frequency of putative splicing enhancers and silencers, consecutive PESEs/PESSs that had less than four differing bases were considered to be part of a single motif. Splice motif frequency was also examined using RESCUE-ESE (Yeo et al. 2005), and both methods produced congruent results.

RT-PCR

RNA was isolated from mouse, rat, chicken, and *Xenopus laevis* whole brain using TRIzol (Invitrogen) according to the manufacturer's instructions. The RNA was reverse transcribed with SuperScript III (Invitrogen). Human and cow brain cDNA were kind gifts from Dr. Peter Dodd (University of Queensland) and Professor Michael D'Occhio (University of Queensland), respectively. cDNA was amplified with species-specific primer sets (Supplemental Table S2). For the primer set targeting A2/B1 exon 9, PCRx Enhancer (Invitrogen) was added to the PCR mix to improve amplification. PCR products were run on 3% ultra-high-resolution agarose gel. Band intensities were calculated using Kodak Imaging Software v 4.5.1 (Kodak). Unknown bands were excised and the DNA was extracted using a gel extraction kit (Promega). The DNA was cloned into pGEMT (Roche) and sequenced to identify novel splice products.

Affinity isolation with bound oligonucleotides and Western blotting

A2RE11- and Telo4-binding rat brain proteins were isolated by affinity pulldowns as previously described (Hoek et al. 1998). Briefly, homogenized P21 rat brain was incubated with Dynabeads (Invitrogen) bearing immobilized biotinylated oligonucleotides for A2RE11 (GCCAAGGAGCC) and Telo4 (TTAGGGTTAGGGTTAGGGTTAGGG) (Geneworks). Bound proteins were eluted in 30% (v/v) acetonitrile in 0.1% trifluoroacetic acid and analyzed by Western blotting with the following rabbit antibodies raised in our laboratory: A1* (1:500), A2/B1 (1:10,000), $A3^{NT}$ (1:10,000), and $A3^{CT}$ (1:500) (Ma et al. 2002), followed by Alexafluor 680 goat anti-rabbit (1:20,000; Invitrogen). Homogenized human neuroblastoma cell line (SH-SY5Y) was also analyzed using $A3^{NT}$ and $A3^{CT}$.

Translation assay

All primers and oligonucleotides used for cloning are listed in Supplemental Table S3. As the sizes of the inserts were very small, the original start codons of the pEGFP vectors were mutated to prevent leaky scanning. mGFP was made by site-directed mutagenesis of pEGFP-N1 (Clontech) with GFPmut1 and GFPmut2. This converted the Kozak consensus sequence and start site from GCCACCATG to GGCGGCGTG. The GeneDimmer expression vectors, pEGFP-2mer and pEGFP-4mer, which contain two and four uORFs, respectively, with a 52-nt spacer between uORFs (Wang and Rothnagel 2004), were mutated to 2mer-mGFP and 4mer-mGFP using GFPmut3 and GFPmut4 (GeneDimmer). This removed the start codon and introduced a BglII site into the vector, which already contained an AgeI site.

A3ex1b and A3ex2 were made by annealing A3ex1bF and A3ex1bR, and A3ex2F and A3ex2R, respectively. These oligonucleotides were designed such that the annealed product would be flanked by a 5′ BglII site and start codon and a 3′ AgeI site with sticky ends. The annealed products were then ligated into digested mGFP constructs.

HeLa cells were grown in six-well plates and transfected with the pEGFP constructs using lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Two micrograms of DNA was added per well and three sets of transfections were performed for each construct. The following day, the intensity of pEGFP expression in the cells was analyzed on a BD FACSAria cell sorter (BD Biosciences), and 50,000 pEGFP-positive cells were counted per construct.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

This work was supported by grants from the Australian National Health and Medical Research Council and the Cancer Council of Queensland to R.S. and J.A.R. S.P.H. is supported by a University of Queensland Research Scholarship. We thank Dr. Markus Kerr for providing the SH-SY5Y cells and Drs. Peter Dodd, Shaniko Shini, Michael D'Oochio, and David Adams for providing the human, chicken, cow, and frog brain samples, respectively.

Received February 26, 2010; accepted June 3, 2010.

REFERENCES

Ainger K, Avossa D, Diana AS, Barry C, Barbarese E, Carson JH. 1997. Transport and localization elements in myelin basic protein mRNA. *J Cell Biol* **138:** 1077–1087.

Akindahunsi AA, Bandiera A, Manzini G. 2005. Vertebrate 2xRBD hnRNP proteins: A comparative analysis of genome, mRNA and protein sequences. *Comput Biol Chem* **29:** 13–23.

Alekseyenko AV, Kim N, Lee CJ. 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* **13:** 661–670.

Ast G. 2004. How did alternative splicing evolve? *Nat Rev Genet* **5:** 773–782.

Bowes JB, Snyder KA, Segerdell E, Jarabek CJ, Azam K, Zorn AM, Vize PD. 2010. Xenbase: Gene expression and improved integration. *Nucleic Acids Res.* **38:** D607–D612.

Buvoli M, Cobianchi F, Bestagno MG, Mangiarotti A, Bassi MT, Biamonti G, Riva S. 1990. Alternative splicing in the human gene for the core protein-A1 generates another hnRNP protein. *EMBO J* **9:** 1229–1235.

Chabot B, Blanchette M, Lapierre I, LaBranche H. 1997. An intron element modulating 5′ splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol* **17:** 1776–1786.

Dreyfuss G, Matunis MJ, Pinolroma S, Burd CG. 1993. hnRNP proteins and the biogenesis of messenger-RNA. *Annu Rev Biochem* **62:** 289–321.

Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3:** 195–205.

Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297:** 1007–1013.

Gao Y, Tatavarty V, Korza G, Levin MK, Carson JH. 2008. Multiplexed dendritic targeting of α-calcium calmodulin-dependent protein kinase II, neurogranin, and activity-regulated cytoskeleton-associated protein RNAs by the A2 pathway. *Mol Biol Cell* **19:** 2311–2327.

Han SP, Friend LR, Carson JH, Korza G, Barbarese E, Maggipinto M, Hatfield JT, Rothnagel JA, Smith R. 2010. Differential subcellular distributions and trafficking functions of hnRNP A2/B1 spliceoforms. *Traffic* **11:** 886–898.

Hatfield JT, Rothnagel JA, Smith R. 2002. Characterization of the mouse hnRNP A2/B1/B0 gene and identification of processed pseudogenes. *Gene* **295:** 33–42.

He YW, Smith R. 2009. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci* **66:** 1239–1256.

He YW, Brown MA, Rothnagel JA, Saunders NA, Smith R. 2005. Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation. *J Cell Sci* **118:** 3173–3183.

Hertel KJ. 2008. Combinatorial control of exon recognition. *J Biol Chem* **283:** 1211–1215.

Hoek KS, Kidd GJ, Carson JH, Smith R. 1998. hnRNP A2 selectively binds the cytoplasmic transport sequence of myelin basic protein mRNA. *Biochemistry* **37:** 7021–7029.

Hubbard TJP, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, et al. 2009. Ensembl 2009. *Nucleic Acids Res* **37:** D690–D697.

Hutchison S, LeBel C, Blanchette M, Chabot B. 2002. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5′ splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* **277:** 29745–29752.

Kamma H, Horiguchi H, Wan LL, Matsui M, Fujiwara M, Fujimoto M, Yazawa T, Dreyfuss G. 1999. Molecular characterization of the hnRNP A2/B1 proteins: Tissue-specific expression and novel isoforms. *Exp Cell Res* **246:** 399–411.

Kamma H, Satoh H, Matusi M, Wu WW, Fujiwara M, Horiguchi H. 2001. Characterization of hnRNP A2 and B1 using monoclonal antibodies: Intracellular distribution and metabolism through cell cycle. *Immunol Lett* **76:** 49–54.

Kim E, Goren A, Ast G. 2008. Alternative splicing: Current perspectives. *Bioessays* **30:** 38–47.

Koren E, Lev-Maor G, Ast G. 2007. The emergence of alternative 3′ and 5′ splice site Exons from constitutive exons. *PLoS Comput Biol* **3:** 895–908.

Kozak M. 1991. Structural features in eukaryotic messenger-RNAs that modulate the initiation of translation. *J Biol Chem* **266:** 19867–19870.

Krecic AM, Swanson MS. 1999. hnRNP complexes: Composition, structure, and function. *Curr Opin Cell Biol* **11:** 363–371.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

Lev-Maor G, Goren A, Sela N, Kim E, Keren H, Doron-Faigenboim A, Leibman-Barak S, Pupko T, Ast G. 2007. The 'alternative' choice of constitutive exons throughout evolution. *PLoS Genet* **3:** 2221–2234.

Ma ASW, Moran-Jones K, Shan J, Munro TP, Snee MJ, Hoek KS, Smith R. 2002. Heterogeneous nuclear ribonucleoprotein A3, a novel RNA trafficking response element-binding protein. *J Biol Chem* **277:** 18010–18020.

Maggipinto M, Rabiner C, Kidd GJ, Hawkins AJ, Smith R, Barbarese E. 2004. Increased expression of the MBP mRNA binding protein HnRNP A2 during oligodendrocyte differentiation. *J Neurosci Res* **75:** 614–623.

Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res* **33:** D54–D58.

Maris C, Dominguez C, Allain FHT. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272:** 2118–2131.

Matsui M, Horiguchi H, Kamma H, Fujiwara M, Ohtsubo R, Ogata T. 2000. Testis- and developmental stage-specific expression of hnRNP A2/B1 splicing isoforms, B0a/b. *Biochim Biophys Acta* **1493:** 33–40.

Mayeda A, Munroe SH, Caceres JF, Krainer AR. 1994. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J* **13:** 5483–5495.

Modrek B, Lee CJ. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* **34:** 177–180.

Moran-Jones K, Wayman L, Kennedy DD, Reddel RR, Sara S, Snee MJ, Smith R. 2005. hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere. *Nucleic Acids Res* **33:** 486–496.

Munro TP, Magee RJ, Kidd GJ, Carson JH, Barbarese E, Smith LM, Smith R. 1999. Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking. *J Biol Chem* **274:** 34389–34395.

Myer VE, Steitz JA. 1995. Isolation and characterization of a novel, low abundance hnRNP protein-A_0. *RNA* **1:** 171–182.

Needleman S, Wunsch CD. 1970. A general method applicable to search for similarities in amino acid sequence of 2 proteins. *J Mol Biol* **48:** 443–453.

Pan Q, Bakowski MA, Morris Q, Zhang W, Frey BJ, Hughes TR, Blencowe BJ. 2005. Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet* **21:** 73–77.

Roca X, Sachidanandam R, Krainer AR. 2005. Determinants of the inherent strength of human 5′ splice sites. *RNA* **11:** 683–698.

Shan JG, Munro TP, Barbarese E, Carson JH, Smith R. 2003. A molecular mechanism for mRNA trafficking in neuronal dendrites. *J Neurosci* **23:** 8859–8866.

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes—sequence statistics and functional implications in gene-expression. *Nucleic Acids Res* **15:** 7155–7174.

Smith R. 2004. Moving molecules: mRNA trafficking in mammalian oligodendrocytes and neurons. *Neuroscientist* **10:** 495–500.

Sorek R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA* **13:** 1603–1608.

Sueoka E, Sueoka N, Iwanaga K, Sato A, Suga K, Hayashi S, Nagasawa K, Nakachi K. 2005. Detection of plasma hnRNP B1 mRNA, a new cancer biomarker, in lung cancer patients by quantitative real-time polymerase chain reaction. *Lung Cancer* **48:** 77–83.

Tanaka E, Fukuda H, Nakashima K, Tsuchiya N, Seimiya H, Nakagama H. 2007. HnRNP A3 binds to and protects mammalian telomeric repeats in vitro. *Biochem Biophys Res Commun* **358:** 608–614.

Venables JP, Koh CS, Froehlich U, Lapointe E, Couture S, Inkel L, Bramard A, Paquet ER, Watier V, Durand M, et al. 2008. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* **28:** 6033–6043.

Wang ZF, Burge CB. 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* **14:** 802–813.

Wang XQ, Rothnagel JA. 2004. 5′-Untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. *Nucleic Acids Res* **32:** 1382–1391.

Wu SL, Sato M, Endo C, Sakurada A, Dong BM, Aikawa H, Chen Y, Okada Y, Matsumura Y, Sueoka E, et al. 2003. hnRNP B1 protein may be a possible prognostic factor in squamous cell carcinoma of the lung. *Lung Cancer* **41:** 179–186.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11:** 377–394.

Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. 2005. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci* **102:** 2850–2855.

Functional implications of the emergence of alternative splicing in hnRNP A/B transcripts

Siew Ping Han, Karin S. Kassahn, Adam Skarshewski, et al.

RNA 2010 16: 1760-1768 originally published online July 22, 2010
Access the most recent version at doi:10.1261/rna.2142810