Noncanonical RNAs From Transcripts of the *Drosophila muscleblind* Gene

JONATHAN M. HOUSELEY*, ZAIDA GARCIA-CASADO*, MAYA PASCUAL, NURIA PARICIO, KEVIN M. C. O'DELL, DARREN G. MONCKTON, AND RUBEN D. ARTERO

From the Department of Genetics, University of Valencia, Doctor Moliner 50, 46100 Burjasot, Valencia, Spain (Garcia-Casado, Pascual, Paricio, and Artero); and Institute of Biomedical and Life Sciences, University of Glasgow, Anderson College Complex, 56 Dumbarton Road, Glasgow G11 6NU, UK (Houseley, O'Dell, and Monckton). Jonathan M. Houseley is now at The Wellcome Trust Centre for Cell Biology, Michael Swann Building, Mayfield Road, Edinburgh EH9 3JR, UK; and Zaida Garcia-Casado is now at the Department of Haematology, Hospital Universitario La Fe, Valencia, Spain.

Address correspondence to R. D. Artero at the address above, or e-mail: ruben.artero@uv.es.
*These authors contributed equally to this work.

Abstract

It has become increasingly evident that eukaryotic cells produce RNA molecules from coding genes with constitutions other than those of typically spliced mRNA transcripts. Here we describe new cDNAs from the *Drosophila melanogaster muscleblind* (*mbl*) locus that identify two such atypical RNA molecules: RNAs containing an incomplete exon 2 tandem repetition (*mblE2E2'*) or having exons with a different order compared to the corresponding genomic DNA (*mblE2E3' E2'*; exon scrambling). The existence of exon duplications and rearrangements in the genomic locus that might explain such cDNAs was ruled out by genomic Southern blotting and in silico analysis of the *Drosophila* genome sequence. The incomplete exon 2 tandem repetition was confirmed by sequencing reverse transcriptase-polymerase chain reaction (RT-PCR) products, rapid amplification of cDNA ends, and detection of a band consistent with cDNA sizes in total RNA northern blots. RT-PCRs with exon-specific primers downstream of exon 2 were unable to amplify products other than those expected from canonical *mbl* isoforms, thus indicating that no other exons were efficiently spliced downstream of exon 2. Moreover, *mblE2E2'* transcripts seem to be poorly polyadenylated, if at all, and behave aberrantly in a polyacrylamide gel electrophoresis (PAGE) mobility assay. Taken together, lack of polyadenylation, lack of downstream splicing events, small size of *mblE2E2'*, and PAGE behavior all suggest that these noncanonical transcripts may be circular RNAs. The functional implications for these noncanonical transcripts are unclear. A developmental expression profile of *mblE2E2'* revealed an almost constant expression except during early embryogenesis and early adulthood. The protein putatively encoded is unlikely to be functional because an in-frame stop codon occurs almost immediately after the splice site. Such noncanonical transcripts have previously been observed in vertebrates, and these data provide the first experimental evidence for similar phenomena in invertebrates.

Transcript heterogeneity is a feature of many genes and one that serves in many cases to facilitate additional functionality. Several mechanisms, which occur at both the DNA and RNA levels, have been shown to contribute to this heterogeneity, all of which involve either the rearrangement of sequences within a genome or the use of alternative signals in contiguous RNA. However, the basic requirement to maintain the linear order of gene sequences in RNA transcripts has been challenged in recent years by the description of nontypical RNA molecules. Some of these molecules differ from canonical mRNAs in having tandem duplications of one or more exons in the context of a typical mRNA in the absence of duplications or rearrangements in the DNA. Such exon repetition events have been proposed to constitute a new

pathway for transcript heterogeneity and not to arise from *trans*-splicing as previously assumed (Rigatti et al. 2004). Through serendipity, only a small number of human and rat genes have been found to contain this specific pattern of repetition, the best-studied examples being the rat *COT* and *Sa* genes (Caudevilla et al. 1998; Frantz et al. 1999; Rigatti et al. 2004). Exon repetition has been shown to be allele specific, thus suggesting that it is a *cis*-acting property of the allele rather than a by-product of the splicing machinery (Rigatti et al. 2004). The detection of a larger protein product from the *COT* gene suggests that in some cases the repetition might have a biological role (Caudevilla et al. 1998).

The linear layout of genomic sequences is also altered in transcripts showing exon scrambling, in which exons are

accurately joined at consensus splice sites but in a different order compared to the corresponding genomic DNA. Exon scrambling was initially correlated with the presence of large introns (Cocquerelle et al. 1992) and exon skipping (Zaphiropoulos 1996, 1997), and the first transcripts described were neither polyadenylated nor expressed at their normal level (Nigro et al. 1991; Zaphiropoulos 1996, 1997). However, later studies reported scrambled transcripts that were polyadenylated, adjacent to small introns, and joined at nonconsensus sites (Caldas et al. 1998; Crawford et al. 1999). Scrambled exons have properties consistent with a circular RNA (Megonigal et al. 2000; Nigro et al. 1991), and RNA molecules composed of circularized single exons, or circons, have indeed been observed at low levels (Schindewolf et al. 1996; Zaphiropoulos 1996, 1997). The functional and evolutionary implications of these nontypical RNAs are largely unknown, so they are typically regarded as erroneous by-products of the splicing machinery. For the most part, scrambled transcripts do not seem to have the potential for encoding functional proteins. Recent computational searches, however, showed that exon scrambling and repetition events may be relatively widespread affecting up to approximately 1% of mammalian genes. In addition, conservation of the phenomenon in one gene in humans and mice suggests a biological function (Dixon et al. 2005).

Primary transcripts arising from the *Drosophila muscleblind* (*mbl*) locus have been described to follow a canonical alternative-splicing pathway, giving rise to four transcript isoforms known as *mblA*, *mblB*, *mblC*, and *mblD*. Each transcript encodes for a nuclear protein isoform with a constant N-terminus encoding one or two Cys$_3$His zinc fingers (Begemann et al. 1997). *mbl* is an essential gene during *Drosophila* embryogenesis, and homozygous mutant *mbl* embryos die inside the chorion as fully formed first-instar larvae, with severe defects in the terminal differentiation of embryonic muscles (Artero et al. 1998). Mosaic analysis has also revealed that *mbl* is required later in development for photoreceptor differentiation (Begemann et al. 1997). Recent studies with one of the human Mbl orthologs the muscle blind-like 1 protein indicate that Mbl proteins are regulators of alternative splicing of pre-mRNAs (Ho et al. 2004). Human Mbl proteins, in addition, play a pivotal role in the pathogenesis of the myotonic dystrophies (for recent reviews, see Pascual et al. 2006; Ranum and Day 2004).

Here we report that *mbl* primary transcripts, in addition to their canonical splicing pathway, give rise to noncanonical RNA molecules. Noncanonical transcripts are small, poorly polyadenylated, strongly retarded in high-percentage polyacrylamide gel electrophoreses (PAGEs), and characterized by a partial exon 2 tandem repetition in the absence of additional exonic sequences. All these properties are consistent with circular RNA molecules. We also describe an example of exon scrambling at nonconsensus splice sites within exon 2 and exon 3. To our knowledge, experimental characterization of such noncanonical RNAs has been limited to vertebrate genes (including humans), making *mbl* the first example of a nontypical *Drosophila* RNA that is verified in vivo.

Materials and Methods

cDNA Isolation and Characterization

A 12- to 24-h embryonic cDNA library from an isogenic second chromosome stock *dp cn b*w (Brown and Kafatos 1988) was screened using a 0.6-kb *Hin*dIII genomic fragment, containing most of the *mbl* exon 2, by following standard procedures adapted to the use of nonradioactive probes (Ausubel et al. 1992; Roche Molecular Biochemicals, Mannheim, Germany). Six cDNAs were isolated, and their ends were manually sequenced with the use of Sequenase (United States Biochemical, Cleveland, OH). Sequence analyses were performed with programs from the GCG bioinformatics package (Genetics Computer Group, University of Wisconsin).

Reverse Transcriptase-Polymerase Chain Reaction Amplification and Rapid Amplification of cDNA Ends

Total RNA from Oregon-R wild-type strain was prepared by the guanidinium isothiocyanate method (TriReagent, Sigma, St. Louis, MO), and contaminating genomic DNA was eliminated by incubating with RNase-free DNaseI for 15 min at 37°C. cDNA was made from 1 μg RNA using the AMV reverse transcriptase (RT) kit (Roche Molecular Biochemicals) following the manufacturer's recommendations for oligo(dT)-primed synthesis. Of the 20 μl cDNA synthesis mixture, 2 μl was directly used for polymerase chain reaction (PCR) amplification. Primers used for PCR amplification were as follows: B, 5′-CCCACAAGCTTTGTAATCC-3′; A, 5′-GCTGATAGCTTGGGCTAT-3′; E3, 5′-TTCCCC-TCGGCGCTTTTC-3′; E4, 5′-GCACGGCGGTTTATC-ACG-3′; and E5a, 5′-GCTGTGCCACCTGCACTG. For the primer combination A and B, cycling conditions were denaturation for 30 s at 94°C, annealing for 30 s at 55°C, and extension for 1 min at 72°C. Thirty amplification cycles were performed in a total volume of 20 μl. For all other primer combinations, the annealing temperature was 50°C.

For rapid amplification of cDNA ends (3′ RACE), RNA isolated as above was reverse transcribed from ADAPT-dT (5′-CGAGGGGGATGGTCGACGGAAGCGACCT-TTTTTTTTTTTTTTT-3′) using SuperScript II (Invitrogen, Carlsbad, CA) and amplified using mblF2 (5′-CTATAA-CCCCAAAGATCTATGGCC-3′) and X45 (5′-CGAGGG-GGATGGTCGACGG-3′), with 35 cycles of 94°C for 45 s, 60°C for 45 s, and 72°C for 3 min. Products were gel purified and cloned into pGEM T-Easy (Promega, Madison, WI) for sequencing. Existence of rearranged exon 2 was assessed in developmentally staged cDNA reverse transcribed from random hexamers with SuperScript II using primers mblF5 (5′-GGATTCGCGCTGGCTGCAATTGG-3′) and mblXR (5′-GCTTGTGGGGGTAGCTGATAGCTTGGG-3′); PCR conditions were as above, except the extension time was reduced to 1 min.

Northern Blot Analysis

RNA samples (1 μg total Oregon-R RNA or 200 ng poly(A)$^+$ RNA isolated with a Dynabeads® mRNA Direct™ kit) were denatured at 85°C for 10 min in 0.2% sodium dodecyl sulfate,
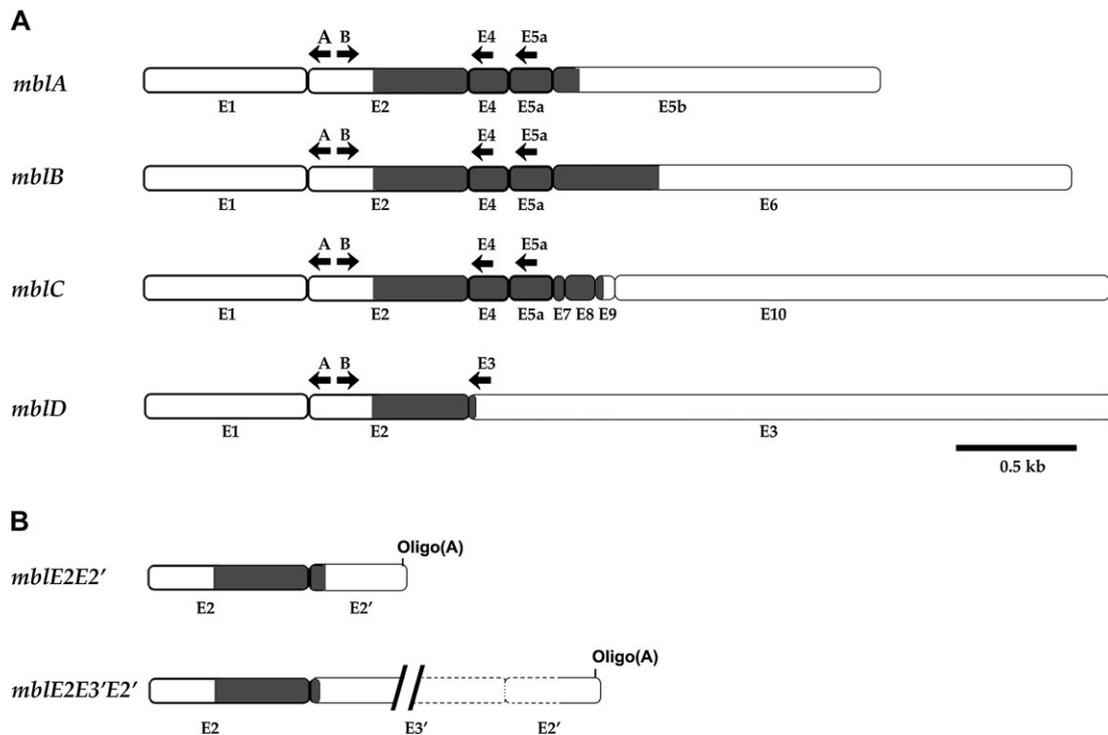
Figure 1. Schematic representation of the canonical and atypical mRNA isoforms expressed by the *Drosophila mbl* gene and of the primers used in the PCR amplifications. (A) Normal transcripts of the *mbl* isoforms *A*, *B*, *C*, and *D* are illustrated, with indication of the exon combination used in each case. Primer names and approximate locations are given. Shaded areas denote coding regions. (B) Representation of the exonic structure of cDNA clones isolated in this work. Clones 2.3, 2.4, and 2.8 represent transcripts with an incomplete exon 2 tandem repetition (*mblE2E2′* transcripts), while clone 2.6 represents an example of exon scrambling (*mblE2E3′ E2′*). Oligo(A) tracts are indicated by a vertical line, and the coding region is shaded. Unknown sequence in *mblE2E3′ E2′* is represented with a dotted line. The figure is drawn to scale.

snap chilled, and separated through a 1.5% tris-acetate EDTA agarose gel with 20 mM guanidine thiocyanate and 200 ng/ml ethidium bromide. The RNA was transferred onto a nylon membrane in 20× standard saline citrate and hybridized with digoxigenin-11-dUTP-labeled antisense riboprobes to *mbl* and *rp49* at 68°C in EasyHyb (Roche Molecular Biochemicals). For PAGE analysis of *mblE2E2′*, 10 μg total Oregon-R RNA and 1 μg New England Biolabs (Beverly, MA) 50-bp ladder were separated in 4%–8% polyacrylamide/7 M urea gels containing 1× tris-borate-EDTA. Gels were briefly stained with approximately 150 ng/ml ethidium bromide and imaged before transfer overnight to HyBond N+ membrane in 0.5× TBE at 8 V. Blots were probed with ^{32}P-labeled, random-primed probes according to standard protocols.

Results

Isolation of Noncanonical cDNA Clones From the *mbl* Locus

The alternative use of the 11 exons described in the *mbl* transcription unit gives rise to four transcript isoforms (Figure 1A; Begemann et al. 1997; Garcia-Casado Z, unpublished observations). To further characterize the alternative splicing of the gene, an additional embryonic cDNA library was screened with a genomic probe containing most of exon 2, which includes the putative start codon of the Mbl protein (see Materials and Methods). Surprisingly, all six cDNAs isolated were small, ranging in size from 0.9 to 1.2 kb, and showed an unusual structure. Restriction analysis and sequencing of cDNA ends revealed that clones 2.3, 2.4, and 2.8 had an incomplete tandem repetition of exon 2. The tandem repetition consisted of a complete exon 2 followed by the acceptor site of another exon 2 to an interrupted adenine-rich region lying approximately 100 bp 5′ of the start codon. We refer to this noncanonical transcript structure as *mblE2E2′*, where E2′ denotes exon 2 sequence from the start of exon 2 to the adenine-rich region. Similar analyses with cDNA 2.6 revealed a structure in which a partial exon 3 and a partial exon 2 (to the adenine-rich region) followed a complete exon 2, an exonic organization we refer to as *mblE2E3′ E2′* (Figure 1B). Rearrangements of exonic sequences such as the one present in cDNA 2.6 have been defined as exon scrambling (Nigro et al. 1991). While the incomplete exon repetition event in *mblE2E2′* used genuine splice sites, we infer that scrambled exons in *mblE2E3′ E2′* used a cryptic donor splice site in E3 because exon E3 is larger (2.6 kb) than the size of cDNA 2.6 (<1.2 kb).

mbl Transcripts With Duplicated Exons Exist In Vivo

A number of possibilities might explain the isolation of atypical cDNA clones from the *mbl* locus, most notably the existence of a duplicated and rearranged extra copy of the gene, a possible *trans*-splicing event that would result in exon repetition and exon scrambling, and the artifactual joining of partial cDNAs. Homology searches of the *Drosophila* genome sequence for near-exact matches of exon 2 returned a single hit, thus indicating that no extra copies of exon 2 exist. This result was confirmed by genomic Southern blot analysis, which detected only a single copy of exon 2 in the genome of *Drosophila* strains Oregon-R and Canton-S (data not shown). Intermolecular splicing of two *mbl* pre-mRNAs is also an unlikely explanation because *trans*-splicing of intact genes has been reported very rarely (Horiuchi et al. 2003). If exon repetition and scrambling resulted from a *trans*-splicing event, it is likely that this would generate two products: a long mRNA having a duplicated exon 2 and a short mRNA lacking that exon and having exon 1 directly joined to exon 3 (or exon 4). No cDNAs corresponding to the short transcript have been isolated. *Trans*-splicing, in addition, would require the use of cryptic splice sites inside exons 2 and 3 in order to explain structures such as *mblE2E2′* and *mblE2E3′ E2′* and would include the remaining exonic sequences to the end of the transcript.

Noncanonical *mbl* transcripts were detected in vivo at several developmental stages by RT-PCR and 3′ RACE. PCR primers in exon 2 that are oriented outward gave a strong product from cDNA but not from genomic DNA, confirming the absence of this arrangement at the DNA level and indicating that the exon 2 repetition takes place during primary *mbl* transcript maturation. The PCR product was purified and sequenced, confirming the repeated structure found in cDNAs 2.3, 2.4, and 2.8 as well as the use of genuine donor and acceptor splice sites (Figure 2A,B,D). Similarly, 3′ RACE with an oligo(dT) primer gave a product whose sequence also confirms both the exon 2 repetition and the use of correct splice sites (Figure 2C). Like cDNAs 2.3, 2.4, and 2.8, the 3′ end of the RACE product is an interrupted adenine-rich region of exon 2 that presumably binds the oligo(dT) cDNA synthesis primer. Transcripts with the structure *mblE2E2′* could encode a small protein with a partial Cys$_3$His zinc finger motif (Figure 2E). However, the presence of an in-frame stop codon near the splice site makes it unlikely that a functional protein is actually translated.

Noncanonical Transcripts From the mbl Locus Are Abundant but Poorly Polyadenylated

The abundance of noncanonical *mbl* transcripts, such as *mblE2E2′*, was investigated by northern blot hybridization of total RNA using an antisense *mblA* probe. Unexpectedly, although no known *mbl* transcript isoforms were observed in this blot, a strong signal was detected at approximately 0.65 kb, which indicates that noncanonical *mbl* transcripts are very common (Figure 3A). Northern blot analysis of poly(A)$^+$ RNA under similar conditions, in contrast, detected a faint band around 0.65 kb, while canonical *mbl* transcripts such

as *mblA*, *mblB*, and *mblC* were clearly detected (Figure 3B). The difference in intensity of the 0.65-kb band in total and poly(A)$^+$ RNA demonstrates that this product is poorly polyadenylated. Densitometric analysis of the poly(A)$^+$ RNA northern revealed that the 0.65-kb species comprises less than 10% of the total polyadenylated *mbl* transcript pool (data not shown). Conversely, the absence of a detectable signal from the known transcripts in total RNA (even after overexposing the 0.65-kb band) suggests that the 0.65 kb is present at steady-state levels at least 10-fold those of the canonical transcripts.

The developmental expression profile of *mblE2E2′*, as an example of noncanonical RNA, was generated using RT-PCR amplification of developmentally staged cDNAs. To specifically amplify cDNA sequences containing a repetition of exon 2, we used the same approach described in Figure 2A, that is, primers oriented outward from within exon 2 so that a product should be amplified from cDNA but not from the genomic DNA. This experiment showed the expression of *mblE2E2′* at approximately constant levels, except that no early embryonic and early adulthood expression was detected (Figure 3C).

mblE2E2′ Transcripts Do Not Include Downstream Exonic Sequences

In order to detect genuine or aberrant splicing of exon 2 sequences to downstream exons in noncanonical transcripts, the B primer was used in combination with exon E3-, E4-, and E5a-specific primers in RT-PCR amplifications. In these experiments, the B oligonucleotide could prime DNA synthesis from annealing sites located either in E2, thus detecting canonical splice events, or in E2′, which would detect additional splicing events downstream of E2′ (note that because the exon 2 repetition is incomplete, the product size arising from both possibilities would be different; see Figure 2B). In these experiments, only the amplification product from canonical transcripts was detected (Figure 4A), indicating that no additional known exonic sequences are spliced 3′ to exon 2 sequences in noncanonical transcripts.

mblE2E2′ Represents Two RNA Populations

To assess whether *mblE2E2′* transcripts are in fact circular, total *Drosophila* RNA was separated through various percentages of denaturing polyacrylamide gel (Figure 4B). The mobility of linear RNA species in polyacrylamide is invariant with gel percentage relative to linear DNA molecular weight markers; however, circular species are strongly retarded by increasing acrylamide concentration. Two species are highlighted by the *mbl* probe: a faster migrating species that is the correct size for *mbl* exon 2 and a slower migrating species that fails to leave the wells in the 6% polyacrylamide (and in 8%, data not shown). The migration of the faster species is invariant with gel percentage compared to the DNA ladder and an *rp49* mRNA control, showing that this species is linear. The slower species migrates differently in the 4% and 6% gels, and the separation between this species and the apparently smaller rRNA is dramatically decreased in the 4% relative to the 6% acrylamide; the opposite would be expected if
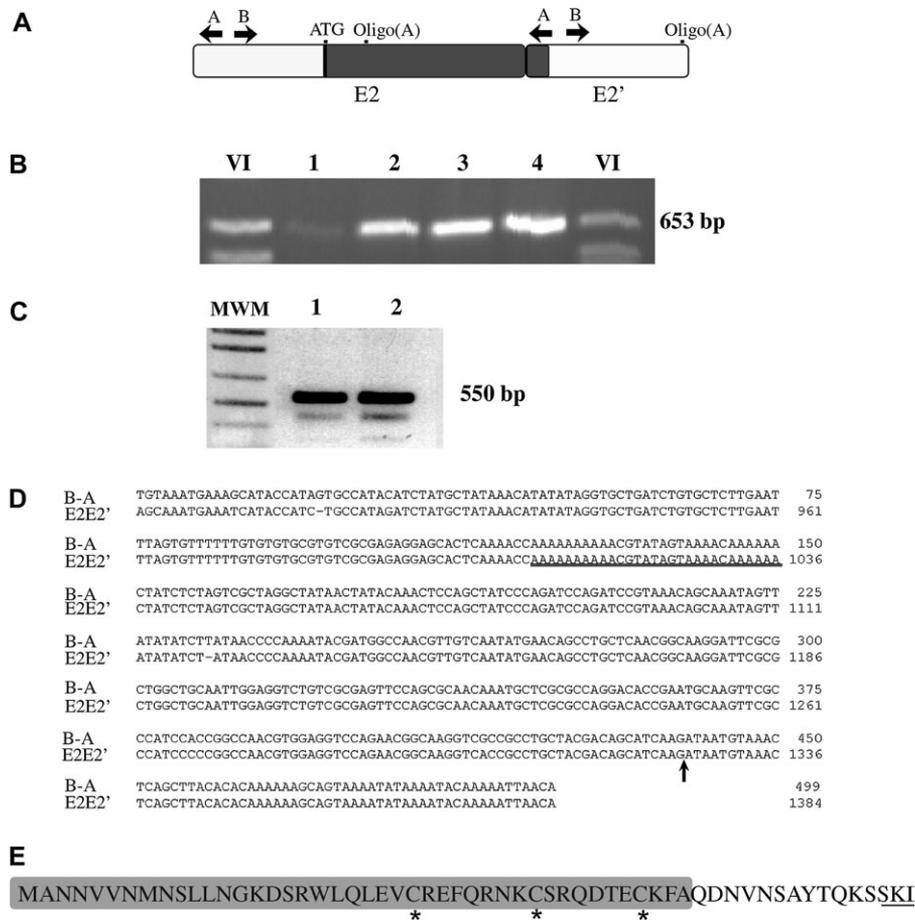
Figure 2. RT-PCR analysis for the occurrence of the exon 2 repetition in vivo and the protein potentially encoded. (A) Schematic representation of an *mblE2E2′* transcript. The presence or absence of exon 2 repetition in *mbl* mRNA was evaluated by RT-PCR amplification with primers A and B. Because both primers face outward, they are unable to amplify any product unless the exon 2 is repeated. The coding region is shaded. E2′ denotes an exon 2 from its acceptor site to the oligo(A) tract. (B) Agarose gel electrophoresis of the RT-PCR products generated with the use of the primer combination A and B. Template RNA was isolated from various developmental stages: lane 1, embryos; lane 2, larvae; lane 3, pupae; and lane 4, adults. Molecular weight marker is number VI (Roche Molecular Biochemicals). (C) 3′RACE analysis of *mbl* transcripts. A 35-cycle PCR was performed with primers mblF2 and X45 on ADAPT-dT–primed cDNA from third-instar larval (lane 1) and adult (lane 2) total RNA. One major product is visible at approximately 550 bp. Molecular weight marker is a 1-kb ladder. (D) Sequencing of amplification product shown in (B) established that it represents an exon 2 duplication in which an exon 2 joined at its donor splice site with another exon 2 at its acceptor splice site. The alignment shows the almost perfect match between the computationally duplicated exon 2 sequence (E2E2′) and the actual sequence (B-A). The splice site is indicated with an arrow, and the oligo(A) tracts are underlined. (E) Transcript of the type *mblE2E2′* could potentially encode a very small protein with a partial zinc finger (conserved cysteines marked with an asterisk) and a microsomal transport signal (underlined).

this band represented a high–molecular weight linear species. We also note the sharpness of the *mbl* band compared to the blurry *rp49* signal, which is consistent with *mblE2E2′* being poorly polyadenylated. In summary, the aberrant behavior of this species in various gel concentrations strongly suggests it is the circular species detected by RT-PCR.

Discussion

In this paper, we describe new noncanonical transcripts from the *mbl* locus characterized by an incomplete repetition of exon 2 and scrambling of exons 2 and 3. We demonstrate that these transcripts are small and abundant, show some degree of developmental regulation, are poorly polyadenylated, and are strongly retarded in high-percentage acrylamide gels. Moreover, we show that no additional exonic sequences are spliced 3′ to the duplicated exon 2. Taken together, the properties of these RNA molecules are consistent with a circular RNA, as similarly proposed for the *ets*-1, *Sry*, or cytochrome *P450 2C24* gene transcripts (Capel et al. 1993; Cocquerelle et al. 1993; Zaphiropoulos 1996). Circular RNAs seem rather stable molecules due to the absence of exonuclease-sensitive
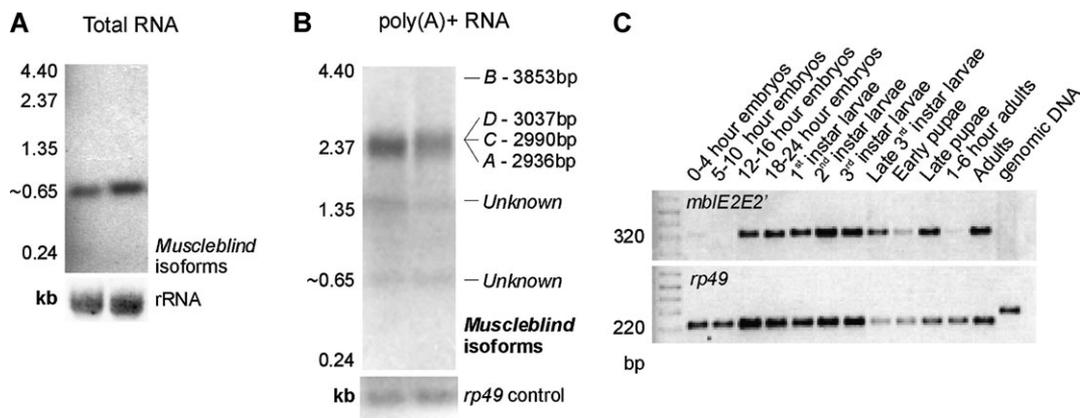
Figure 3. Noncanonical *mbl* transcripts are developmentally regulated and seem to lack a poly(A) tail. (A) Northern blot analysis of *mbl* in third-instar larval total RNA. Total RNA (1 μg) was separated on a 1.5% guanidine thiocyanate agarose gel and probed with DIG-labeled antisense *mblA* RNA. rRNA is ribosomal RNA visualized by ethidium bromide staining as a loading control. Only one RNA isoform of approximately 650 nt is visible. No signal from known *mbl* transcripts was visible on prolonged exposure. Independent extractions are loaded in each lane. (B) Northern blot analysis of *mbl* in third-instar larval mRNA. Poly(A)$^+$ RNA (200 ng) was separated on a 1.5% guanidine thiocyanate agarose gel and probed with DIG-labeled antisense *mblA* RNA followed by antisense *rp49* RNA as a loading control. Independent extractions were used in each lane. Note the faint signal of noncanonical *mbl* transcripts compared to (A). (C) RT-PCR amplification of *mblE2E2′* at the indicated developmental stages. The amplification product from genomic DNA is shown in the right lane and is larger than the cDNA product for *rp49* because the primers span an intron.

5′ and 3′ ends (Cocquerelle et al. 1993). Therefore, even low levels of RNA processing into noncanonical transcripts might lead to the observed accumulation in the *Drosophila* cells (Figure 3).

In a number of cases, exon scrambling and exon skipping have been shown to be reciprocal events, thus suggesting a mechanistic relationship (Zaphiropoulos 1996, 1997). However, the study of other gene transcripts such as *MLL* (Caldas et al. 1998) and our own results with *mbl* transcripts do not support a correlation. In particular, all *mbl* transcript isoforms described to date contain exon 2, which also harbors the start codon of the encoded protein, making it unlikely that an exon 2–skipping event was the origin of the noncanonical transcripts we describe. Large introns adjacent to the exons that are scrambled, as well as the presence of other structural characteristics in the pre-mRNA such as large inverted repeats that would facilitate or stabilize the formation of an intermediary stem-loop, have been proposed to influence exon scrambling (Capel et al. 1993; Cocquerelle et al. 1992). Both requirements are fulfilled in *mbl* primary transcripts: First, the introns between exons 2 and 3 and between exons 3 and 4 are approximately 15 and 60 kb long, respectively, which is relatively large for a *Drosophila* gene. Second, the intronic sequence between exons 2 and 3 contains the non-LTR retrotransposon *jockey J-1* immediately upstream of exon 2, while another non-LTR retrotransposon, an *F element*, is found downstream of exon 3. The presence of partially homologous sequences between both elements could facilitate the hairpin formation in primary transcripts, which is an intermediary potentially implicated in the mechanism of exon scrambling.

Exon 2 of *mbl* transcripts contains a number of oligo(A) tracts upstream of the start codon. In a recent study, the mo-lecular basis for the allele-specific exon repetition event associated with the *Sa* gene (Rigatti et al. 2004) was concluded to involve *cis*-acting elements. Among the potentially involved sequences were six single nucleotide polymorphisms and a difference of five nucleotides in the length of an oligo(A) tract. It is thus temping to propose that tracts of oligo(A) may participate in the processing or regulation of noncanonical transcripts.

Evidence presented herein indicates unusual features of the processing of *mbl* primary transcripts, possibly leading to the generation of circular RNA molecules. A basic question about this phenomenon is whether it is unique to *mbl* or a general occurrence to *Drosophila* genes. Recent computational searches identified 245 genes in mammals, and five *Drosophila* genes, showing nonlinear exon splicing, which include exon repetition and exon-scrambling events (Dixon et al. 2005). Interestingly, the same exon 2 repetition event we describe was found in four expressed sequence tags from the *mbl* gene, thus providing independent confirmation of our results. Additionally, nonlinear exon splicing was identified in another four *Drosophila* genes: *capulet*, *G protein γ30A*, *centaurin gamma1A*, and *CG32306*. These observations suggest that the exon repetition and exon-scrambling phenomena is not limited to *mbl* but is nonetheless limited to a relatively small subset of specific genes. However, it should be considered that the libraries employed to date to search for such transcripts effectively select against circular (poorly polyadenylated) transcripts, and nonlinear exon splicing is thus likely to be more common than currently appreciated (Dixon et al. 2005).

The splicing mechanism yielding circular RNAs may simply represent errors in the normal splicing process, although a possible role of the scrambled transcripts as a feedback regulator of alternative pre-mRNA splicing cannot be
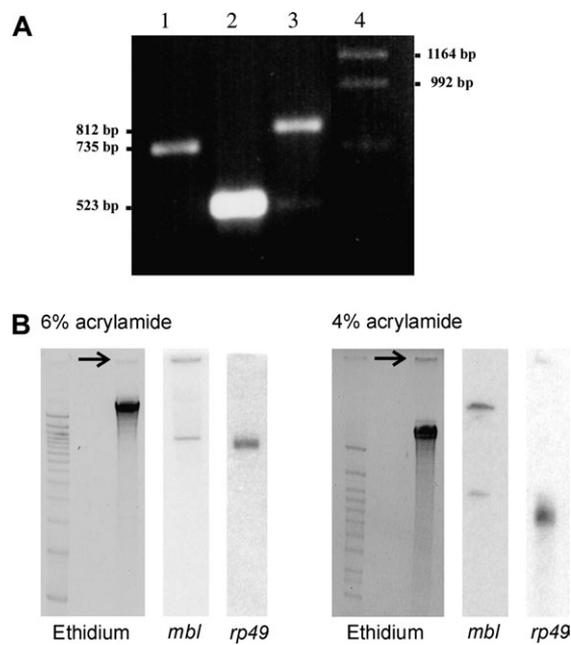
Figure 4. Noncanonical transcripts lack exonic sequences downstream of exon 2 and behave as circular molecules. (A) The result of RT-PCR amplification from 0- to 24-h embryonic total RNA using primer B in combination with primers E3 (lane 1), E4 (lane 2), and E5a (lane 3) is shown. The sizes of the amplification product, if a canonical cDNA transcript isoform was used as template, are 735 bp, 523 bp, and 812 bp, respectively. In all lanes, the size of the amplification product is very close to the expected size, thus indicating that no additional exonic sequences downstream of E2′ exist. Lane 4 includes molecular weight marker VI (Roche Molecular Biochemals). (B) PAGE analysis of *mblE2E2′*. Denaturing polyacrylamide gel analysis of *Drosophila* total RNA. Total RNA (10 μg) was separated on 4% and 6% polyacrylamide gels, along with a linear DNA marker. Gels were stained with ethidium bromide prior to transfer, and membranes were probed sequentially with *mbl* 3′ RACE product and *rp49* as marked. Arrows indicate material remaining in wells. The major signal in the ethidium bromide–stained total RNA represents rRNAs. As discussed in the text, we interpret the blurry *rp49* signal as heterogeneity in poly(A) tail length, while the sharp *mbl* band is consistent with a poorly polyadenylated transcript.

excluded. This is especially tempting in the case of Mbl, which itself is involved in the regulation of alternative splicing (Ho et al. 2004). Whatever the answer, occasional errors may have profound impact in cells, such as mutations generated during DNA replication or oncogene transduction by retrovirus, and potential applications can be derived from them.

Acknowledgments

References

Artero R, Prokop A, Paricio N, Begemann G, Pueyo I, Mlodzik M, Perez-Alonso M, and Baylies MK, 1998. The *muscleblind* gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2. Dev Biol 195:131–143.

Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, and Struhl K, 1992. Current protocols in molecular biology. New York: Greene Publishing Associates.

Begemann G, Paricio N, Artero R, Kiss I, Perez-Alonso M, and Mlodzik M, 1997. *muscleblind*, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development 124:4321–4331.

Brown NH and Kafatos FC, 1988. Functional cDNA libraries from *Drosophila* embryos. J Mol Biol 203:425–437.

Caldas C, So CW, MacGregor A, Ford AM, McDonald B, Chan LC, and Wiedemann LM, 1998. Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. Gene 208:167–176.

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, and Lovell-Badge R, 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell 73:1019–1030.

Caudevilla C, Serra D, Miliar A, Codony C, Asins G, Bach M, and Hegardt FG, 1998. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. Proc Natl Acad Sci USA 95:12185–12190.

Cocquerelle C, Daubersies P, Majerus MA, Kerckaert JP, and Bailleul B, 1992. Splicing with inverted order of exons occurs proximal to large introns. EMBO J 11:1095–1098.

Cocquerelle C, Mascrez B, Hetuin D, and Bailleul B, 1993. Mis-splicing yields circular RNA molecules. FASEB J 7:155–160.

Crawford J, Ianzano L, Savino M, Whitmore S, Cleton-Jansen AM, Settasatian C, d'apolito M, Seshadri R, Pronk JC, Auerbach AD, Verlander PC, Mathew CG, Tipping AJ, Doggett NA, Zelante L, Callen DF, and Savoia A, 1999. The PISSLRE gene: structure, exon skipping, and exclusion as tumor suppressor in breast cancer. Genomics 56:90–97.

Dixon RJ, Eperon IC, Hall L, and Samani NJ, 2005. A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species. Nucleic Acids Res 33:5904–5913.

Frantz SA, Thiara AS, Lodwick D, Ng LL, Eperon IC, and Samani NJ, 1999. Exon repetition in mRNA. Proc Natl Acad Sci USA 96:5400–5405.

Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, and Cooper TA, 2004. Muscleblind proteins regulate alternative splicing. EMBO J 23:3103–3112.

Horiuchi T, Giniger E, and Aigaki T, 2003. Alternative trans-splicing of constant and variable exons of a *Drosophila* axon guidance gene, *lola*. Genes Dev 17:2496–2501.

Megonigal MD, Rappaport EF, Wilson RB, Jones DH, Whitlock JA, Ortega JA, Slater DJ, Nowell PC, and Felix CA, 2000. Panhandle PCR for cDNA: a rapid method for isolation of MLL fusion transcripts involving unknown partner genes. Proc Natl Acad Sci USA 97:9597–9602.

Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, and Vogelstein B, 1991. Scrambled exons. Cell 64:607–613.

Pascual M, Vicente M, Monferrer L, and Artero R, 2006. The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. Differentiation 74:65–80.

Ranum LPW and Day JW, 2004. Pathogenic RNA repeats: an expanding role in genetic disease. Trends Genet 20:506–512.

Rigatti R, Jia JH, Samani NJ, and Eperon IC, 2004. Exon repetition: a major pathway for processing mRNA of some genes is allele-specific. Nucleic Acids Res 32:441–446.

Schindewolf C, Braun S, and Domdey H, 1996. In vitro generation of a circular exon from a linear pre-mRNA transcript. Nucleic Acids Res 24: 1260–1266.

Zaphiropoulos PG, 1996. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. Proc Natl Acad Sci USA 93:6536–6541.

Zaphiropoulos PG, 1997. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. Mol Cell Biol 17:2985–2993.