Transgene analysis proves mRNA trans-splicing at the complex *mod(mdg4)* locus in *Drosophila*

Rainer Dorn*, Gunter Reuter, and Andrea Loewendorf

Institute of Genetics, Biologicum, Martin Luther University, D-06120 Halle, Germany

The *Drosophila* BTB domain containing gene *mod(mdg4)* produces a large number of protein isoforms combining a common N-terminal region of 402 aa with different C termini. We have deduced the genomic structure of this complex locus and found that at least seven of the *mod(mdg4)* isoforms are encoded on both of its antiparallel DNA strands, suggesting the generation of mature mRNAs by trans-splicing. In transgenic assays, we demonstrate the ability of *Drosophila* to produce *mod(mdg4)* mRNAs by trans-splicing of pre-mRNAs generated from transgenes inserted at distant chromosomal positions. Furthermore, evidence is presented for occurring of trans-splicing of *mod(mdg4)*-specific exons encoded by the parallel DNA strand. The *mod(mdg4)* locus represents a new type of comlex gene structure in which genetic complexity is resolved by extensive trans-splicing, giving important implications for genome sequencing projects. Demonstration of naturally occurring trans-splicing in the model organism *Drosophila* opens new experimental approaches toward an analysis of the underlying mechanisms.

Trans-splicing is a process that produces mature transcripts by combining exons of independent pre-mRNA molecules and was first reported in trypanosomes (1–3). The capability of mammalian cells for mRNA trans-splicing was first shown by Eul *et al.* (4). Recently, additional reports demonstrated trans-splicing in higher eukaryotes. Li *et al.* (5) identified a splice variant of the human acyltransferase ACAT-1 that contains a 5′-untranslated exon encoded from chromosome 7 connected with exons 1–16 encoded from chromosome 1. In rat liver cells, carnitine octanoyltransferase mRNA variants with duplications of coding exons 2 and 3 not present in the genomic complement have been identified (6). In this case, different proteins are encoded by the splice variants. However, the physiological significance of these trans-splicing events remains largely unknown.

So far no experimental proof for naturally occurring trans-splicing in insects has been documented. Recently, Labrador *et al.* (7) have shown that the *mod(mdg4)-67.2* isoform in *Drosophila* is encoded on both DNA strands and suggested formation of the mature mRNA by trans-splicing of two independent transcripts.

Sequence analysis of a large number of *mod(mdg4)* cDNA clones (8–12) and the available genomic sequence of the *mod(mdg4)* locus (13) indicated that usage of both DNA strands as coding strands is a general property of the *mod(mdg4)* locus. The common exons 1–4, which are found in all identified transcripts, are located at the 5′ end of the locus whereas the alternatively spliced 3′-specific exons are organized in five groups on both DNA strands. In transgenic assays, we provide evidence that trans-splicing is the basic mechanism responsible for production of multiple *mod(mdg4)* isoforms.

Materials and Methods

Constructs and P Element-Mediated Germline Transformation.
The transgene containing the tagged specific exons *mod(mdg4)-55.1* and *mod(mdg4)-53.1* has been constructed as follows. An enhanced green fluorescent protein (EGFP) (CLONTECH) coding sequence (tag1) has been inserted upstream the stop codon of *mod(mdg4)-55.1* and the Flag tag sequence (tag2) upstream the stop codon of *mod(mdg4)-53.1* via introducing restriction sites by PCR. The final construct containing 2-kb upstream of the 5′ splice site of *mod(mdg4)-55.1* exon 5 and 1.2-kb upstream of *mod(mdg4)-53.1* exon 5 was cloned as a 5.2-kb *Not*I fragment into the upstream activating sequence containing transformation vector pUAST (14) in both orientations. The construct was partially sequenced and used for germline mediated transformation. However, by sequencing the transgene specific reverse transcription (RT)-PCR product, we found a frame shift mutation at amino acid position 499 in *mod(mdg4)-55.1*, preventing the expression of the EGFP-fusion protein.

Germline transformation was performed as described in Rubin and Spradling (15) with the w^{1118} strain as a host. Four transgenic lines have been obtained for the construct [pUAST-55.1-tag1-tag2–53.1], two lines on the second chromosome (TG-1 and -2) and two lines on the third chromosome (TG-3 and -4). In the case of the [pUAST-53.1-tag2-tag1–55.1] construct, we obtained two independent lines, both on the third chromosome, TG-A and B). As the GAL4-driver strain, we used the third chromosomal nanos-GAL4-driver line 4937 (Stock Center, Bloomington, IN).

Reverse Transcription–PCR (RT-PCR).
For transgenic assays, total RNA was isolated from females of the indicated genotypes with Trizol (Invitrogen). Total RNA (1 μg) was used for RT with Moloney murine leukemia virus reverse transcriptase and random hexamer primer (Promega) according to the manufacturer's protocol. Usually one-fifth of the product was used for PCR amplification. PCR was performed in a volume of 30 μl with 8 pmol of each primer, 0.15 mM dNTPs, and 1.65 mM $MgCl_2$. For PCR conditions, we used 95°C for 3 min, 55°C for 40 sec, and 72°C for 40 sec for 35 cycles. The reaction (5–10 μl) was separated in agarose gels. In all experiments, the forward primer E4-F (5′-CGCAAATGTTATGGACCCTCTC-3′) annealing to exon 4 was used in combination with specific backward primers: tag1-back, used for specific detection of tagged exon 5 of *mod(mdg4)-55.1* (5′-CTTGTGGCCGTTTACGTC-3′), tag2-back, used for specific detection of tagged exon 5 of *mod(mdg4)-53.1* (5′-CGTCATCGTCCTTGTAGTC-3′), and mod(mdg4)-55.1-RT-B (5′-CCCGTCCTGTTCTTTTTGAGG-3′).

For RT-PCR experiments (shown in Fig. 3), 1 μg of poly(A)$^+$ RNA isolated from 0- to 3-h embryos of *Drosophila melanogaster* was used. The specific backward primers are mod(mdg4)-64.2-RT-B (5′-CAACTTGCAGTCCTTGCCGTC-3′), -53.1-RT-B (5′-GGGTTGGCTGGAAAATTGATTG-3′), -55.3-RT-B (5′-CCAAGGCATCTTTAGGCTTTGTTG-3′), -67.2-RT-B (5′-

```
                  10        20        30        40
53.4  ------------------------------QEKQFEYVVSQKGHVLLL
58.8  ---AKFDYQISVDVGEATMQLANASSAGVVNSNSPFFIVSKYGTKQIM
54.7  EQEDDFKLHLPLLVTRR-KKTPGGSRKQSFDHLEVSFTRSNRGNNLLT
53.6  ------------------------------ACYQLVPNRRGGKNLI
52.2  --------------------------AKDQNKGVLLKRTAQG-EFLV
                                            F/L   G     L

            50        60        70        80        90
53.4  HKKFPFIREKCINGKTYWRCTQYTTK-TKCHGRLHV----LNGKIVHI
58.8  LKQHTFNRHICRDDVTYWRCSQFAV--LRCRARLKTK---LDTLTILN
54.7  IDGKPFTLNRRIKDVCYWECVKLRCKYIKCSARVVTK---SNRISALS
53.6  FQGHMYSVERKYRNSINWVCSKNSNSVLRCPARCVTNPESGNGIKLSH
52.2  VNGKSYKKTRAMQYRTYFHCLT-RN----CPTYYVLV--ELSRRPRLT
            F/Y         W  C          C

          100       110       120       130       140
53.4  KTHNHSPLDQERKQYMKLQLNNV.
58.8  SEHNHEVITKARKYGSLKRQRAEAEAAARAERRQDPLETAATSAPATTT.
54.7  GLHNHP.
53.6  RRHNHPADAFKPHKRCRKRPGDRK.
52.2  RHHEHAQHCLQCY.
            H:H
```
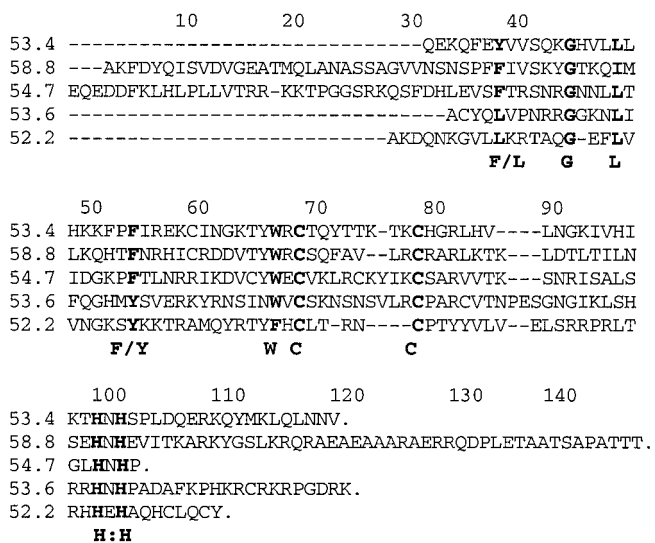
Fig. 1. Sequence comparison of new *Mod(mdg4)* protein isoforms. All protein sequences start with amino acid position 403. The most N-terminal 402 aa common to all *Mod(mdg4)* isoforms are not listed. The conserved C-terminal C_2H_2 consensus sequence is present in all newly identified isoforms. Strongly conserved amino acid positions are shown in bold. Alignment was performed with the CLUSTAL X program (16).

ATATGACTCCCGATTCGCCAGG-3′), -57.4-RT-B (5′-TTGGCCGCCTCAATACGC-3′), and -51.4-RT-B (5′-CAAGACCAATAAGTTTTCAATCCCG-3′).

Isolation and Characterization of cDNA Clones. Two new *mod(mdg4)* cDNA clones [*mod(mdg4)-52.2* and *mod(mdg4)-54.1*] have been isolated from a 2- to 12-h embryonic cDNA library in λZAPII (Stratagene), as described previously (11). The 0.5-kb genomic *Sal*I fragment contained in exon 4 was used as a radioactively labeled probe. The two transcripts *mod(mdg4)-53.6* and *mod-(mdg4)-54.7* have been identified by RT-PCR on poly(A)⁺ RNA

from 0- to 3-h embryos with the primer E4-F and the specific primers mod(mdg4)-53.6–1-RT-B (5′-GGCTTAAAGGCATC-CGCCGGATG-3′) and mod(mdg4)-54.7-RT-B (5′-GGATTT-GGTCACCACACGGGCGGAGC-3′). The missing 3′ end of isoform mod(mdg4)58.8 was identified by RT-PCR with primer E4-F and primer 58.8-RT-B (5′-TACACTTGGACTTAAGAC-TAG-3′).

Results

All *mod(mdg4)* Isoforms Contain Identical N Termini but Variable C Termini. During the molecular analysis of the *mod(mdg4)* locus, we identified 26 different classes of transcripts all containing a common 5′ sequence (exons 1–4) but different 3′ regions. The deduced proteins contain an N-terminal (BTB) domain (17). Furthermore, in most of the isoforms a conserved C-terminal C_2H_2-containing protein motif is found (11). Two new cDNA clones representing isoforms *mod(mdg4)-52.2* and *mod(mdg4)-54.1* were isolated by screening an embryonic cDNA library (see *Materials and Methods*). Two additional putative isoforms, *mod-(mdg4)-53.6* and *mod(mdg4)-54.7* have been detected by searching the genomic region of mod(mdg4) for ORFs containing the C-terminal C_2H_2 consensus sequence. *Mod(mdg4)-58.8* represents another new isoform, which was so far identified by a 3′-truncated cDNA clone. RT-PCR experiments revealed the existence of *mod(mdg4)-53.6*, *-54.7*, and *-58.8* transcripts in early embryos. By sequencing the resulting PCR products, we deduced the putative proteins (Fig. 1). All of them contain the conserved C-terminal protein consensus sequence. Altogether 20 of 26 identified *mod(mdg4)* isoforms combine both the N-terminal BTB domain and the C-terminal consensus sequence, which might be of functional significance.

Genomic Structure of the *mod(mdg4)* Locus. On the basis of our extensive cDNA sequence data and the available sequence of the *mod(mdg4)* region, we deduced the exon/intron structure of *mod(mdg4)* (Fig. 2). Interestingly, seven of the transcripts are not colinearly located within the locus. Relative to the common exons 1–4, exon 5 of isoforms *mod(mdg4)-53.1, -62.3, -55.6, -53.6, -54.7, -57.4,* and *-67.2* are encoded by the antiparallel DNA
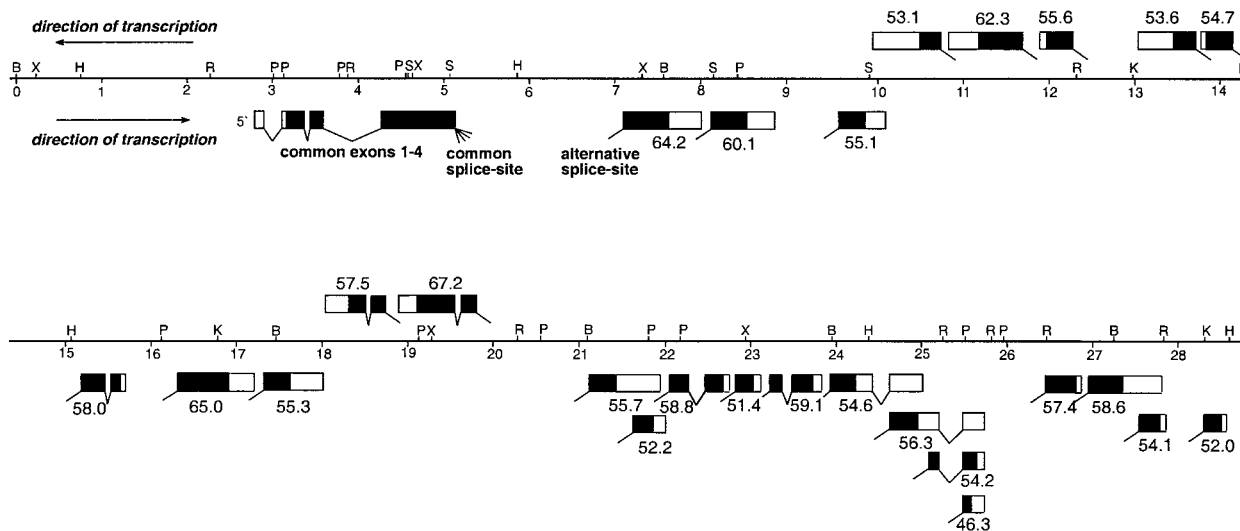

Fig. 2. Genomic region of the *mod(mdg4)* complex and the exon/intron structure of identified mRNAs. The alternative 3′-splice site at the exon 4 boundary is used to generate all mature mRNAs indicated by the molecular weight of the deduced proteins. Exons shown below the physical map are encoded by the same DNA strand, whereas those encoded by the antiparallel DNA strand are shown above the physical map. Direction of transcription is indicated by arrows. Untranslated regions are shown as open boxes and translated regions as black boxes. Note the overlap of specific exons five of isoforms *mod(mdg4)-55.1* and *-53.1* within their 3′-untranslated region and the differential splicing within the specific exons 5 of *mod(mdg4)-55.7* and *-52.2; -54.6, -56.3, -54.2* and *-46.3; -58.6* and *-54.1*. Restriction sites indicated in the physical map are B, *Bam*HI; X, *Xho*I; H, *Hind*III; R, *Eco*RI; P, *Pst*I; S, *Sal*I; and K, *Kpn*I.
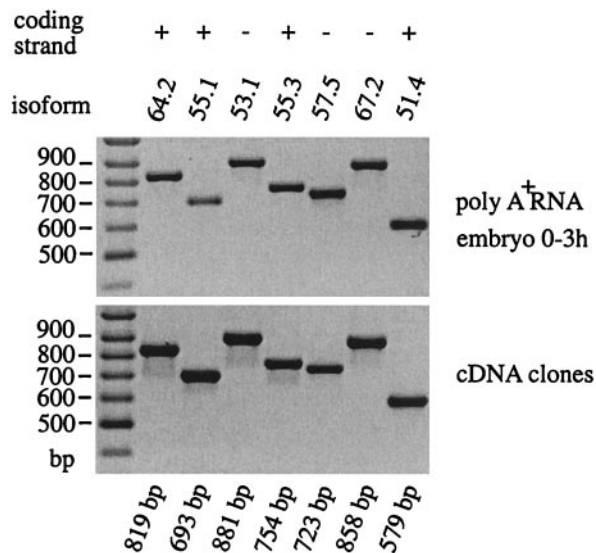
Fig. 3. Results of RT-PCR experiments demonstrating the existence of selected mature mRNAs in poly(A)$^+$ RNA isolated from early embryos (0–3 h). A primer deduced from common exon 4 (E4-F) was used as forward primer and specific backward primers deduced from the alternative exons 5 indicated above the *Upper* panel were used for PCR. Specific exons encoded by the same DNA strand as common exons 1–4 are indicated by (+) and those encoded by the antiparallel DNA strand are indicated by (−). PCR-fragments obtained with the same pairs of primers and the corresponding cDNA clones as DNA template are shown in the *Lower* panel. The expected fragment size for each isoform is indicated at the bottom.

strand. Beside this finding, the exon/intron structure of *mod(mdg4)* suggests differential splicing within isoform-specific exons (isoforms *mod(mdg4)-55.7* and *-52.2*; *mod(mdg4)-54.6*, *-56.3*, *-54.2*, and *-46.3*; *mod(mdg4)-58.6* and *-54.1*). According to its genomic structure, the genetic density at the mod(mdg4) complex is unusually high. Altogether 26 independent transcripts of an average size of 2 kb are encoded in a genomic region of 28 kb.

For most of the mRNAs, more than two independent cDNA clones have been isolated (11). In addition we proved the existence of mRNAs in early embryos for a subset of seven isoforms by RT-PCR. Representative data with the forward primer located within the common exon 4 (primer E4-F) and specific backward primers located within the appropriate specific exons of isoforms *mod(mdg4)-64.2*, *-55.1*, *-53.1*, *-55.3*, *-57.4*, *-67.2*, and *-51.4* are shown in Fig. 3. All amplicons correspond to the expected size deduced from the cDNA clones and sequencing of the RT-PCR products confirmed the sequence of corresponding cDNA clones. These results support the generation of independent mRNAs from both antiparallel DNA strands in case of isoforms *mod(mdg4)-53.1*, *-62.3*, *-55.6*, *-53.6*, *-54.7*, *-57.4*, and *-67.2*. This is in agreement with RNase protection experiments performed on isoform *mod(mdg4)-67.2* (ref. 7).

Furthermore, we compared all 3′ intron sequences upstream of the specific exons 5. In all cases, the conserved sequence elements known to be involved in mRNA splicing, the putative branch point, a pyrimidine-rich tract, and the dinucleotide AG at the 3′ intron border are present (Fig. 4). These data do not indicate significant sequence differences allowing the discrimination between putative cis-splicing of specific exons located colinearly to exons 1–4 and those 3′ exons encoded by the antiparallel DNA strand, which are expected to be trans-spliced.

Specific Exons Transcribed from Transgenes at Different Chromosomes Are Templates for Trans-Splicing. To demonstrate trans-splicing, we have separated specific exons from the common

exons 1–4 to different chromosomes via transgene insertions (Fig. 5). In this assay, we have chosen the specific exons *mod(mdg4)-55.1* and *mod(mdg4)-53.1* (cf. Fig. 1), which are encoded by antiparallel DNA strands. Notably, the 3′-untranslated regions of both isoforms are antisense within a region of 149 nt. This experiment should also test for putative trans-splicing of isoform *mod(mdg4)-55.1*, which is colinearly located with respect to the common exons 1–4. In the transgenes, both specific exons have been sequence-tagged via PCR and the resulting fragment containing genomic sequences of 2.0-kb and 1.2-kb upstream of the splice sites of *mod(mdg4)-55.1* and *mod(mdg4)-53.1*, respectively, was cloned in both orientations into the *Drosophila* transformation vector pUAST (14). Expression of the inserted sequences is induced by the yeast transcriptional activator GAL4, expressed from an independent driver element.

Four transformants containing the tagged specific exon 5 of *mod(mdg4)-55.1* (transgenes TG-1 to 4) and two transformants carrying the tagged specific exon 5 of *mod(mdg4)-53.1* (transgenes TG-A and B) under control of the upstream activating sequence (UAS) element have been obtained. For expression of the transgene, all transformants have been crossed to GAL4-driver lines expressing GAL4 either under hsp70 control or the promoter of the maternal gene nanos. To test for trans-splicing events, total RNA was isolated from progeny females containing the GAL4-driver element and the transgene construct and from females containing the GAL4-driver element but no transgene, respectively. RT-PCR was performed with the forward primer E4-F, which is complementary to common exon 4, and backward primers specific for the two sequence tags. Fig. 6 represents the RT-PCR results obtained from crosses with the nanos driver element. In all females containing the transgene construct and the driver element, an amplicon of the expected size is found, whereas in females not containing the transgene construct no fragment was amplified. Fig. 6*A* represents the results obtained for transgenic lines TG-1 to 4 containing the sequence tagged exon 5 of isoform *mod(mdg4)-55.1* under control of UAS/GAL4. RT-PCR results with the transgenic lines expressing the sequence tagged exon *mod(mdg4)-53.1* under control of GAL4 are shown in Fig. 6*B*. As a control for all genotypes, RT-PCR was performed with the forward primer E4-F and the backward primer 55.1-RT-B annealing upstream to the sequence tag of isoform *mod(mdg4)-55.1* (Fig. 6 *A* and *B Lower*). The expected control PCR fragment of 693 bp was obtained in all genotypes.

Sequencing of the RT-PCR fragments proved exact trans-splicing of endogenous common exon 4 to the tagged specific exon 5 transcribed from the transgene. In the case of the second chromosomal transgenic lines TG-1 and 2, the results demonstrate trans-splicing of mRNAs generated from different chromosomes. Our results furthermore show that the specific exon *mod(mdg4)-55.1*, which is colinearly located with respect to exons 1–4 within the endogenous locus, is trans-spliced in our assay. The latter result also suggests that 3′-specific exons on both strands are joined to the common exons 1–4 by trans-splicing.

Detection of Isoform-Specific Promoter Elements. One of the important criteria for efficient mRNA trans-splicing is the presence of a 3′ splice site in absence of a functional 5′ splice site as represented in outrons (2). To meet this criteria, we would expect the production of independent transcript(s) containing one or several of the endogenous-specific *mod(mdg4)* exons. Searching for putative promoter elements within the *mod(mdg4)* complex, several TATA-box-containing elements were found. One of these is located upstream of the specific exon *mod(mdg4)-55.1* and is contained within the transgene construct used for the trans-splicing assay. To prove its function *in vivo*, all transgenic lines have been tested for expression of the transgene in absence

isoform	coding strand	upstream splice sequence	5`exon sequence
64.2	+	TGCCCACTA**a**ACCCCATG<u>CTTTCGGTTTTCCTCGA</u><u>CTCTC</u>GCAG	ATACGGAGATCAGTT
60.2	+	AGTGTAGACAACTGGAACTAACGATAAA**a**<u>CTTCTTCTTTTC</u>AG	AGGACAGAGCTGGTTT
55.1	+	AATCATTACCATATA**a**TATA<u>TCTTTCCTCCTCTT</u>AAA<u>CTT</u>ACG	GTTTGATTTTTAAGG
53.1	−	CCGACTTACTAACGCACCGGGTCCA**a**TA<u>TCTATTCTCTTTT</u>AAG	GTGACGCCACCCAGT
62.3	−	ATGACCATTTACTAATTACA**a**TGTG<u>TTCCTTTCTCCTTCC</u>ACAG	AAGCAGCGTGGCGAT
55.6	−	CTACGACGAATTCCATTTTTAATTCTA**a**<u>ACCGTTTTATTT</u>ACAG	GACATCTTTCGACAT
58.0	+	TTTGTTTATGTTTTA**a**TCG<u>TTTTTGCCACC</u>GTGTG<u>CCGTTCC</u>AG	ATGGTCCAAGCAAGG
65.0	+	TTAAATCCCAACTAATCTGTTA**a**<u>TCTTTTCATTTT</u>GCAGTCAAG	ATCAAAATGGAGCCT
55.3	+	TTACCAAGACA**a**TTTACTGACTTGATG<u>CTTTTTGTTTCTTT</u>CAG	ATCGCAAACGTTACT
57.5	−	CCCATTTAAATGATAAATAACATA**a**<u>GCCCTTTCATGTCTTT</u>CAG	TGACCCAACACGTCC
67.2	−	ACCCAGATGTTCTA**a**CCTGTTGCG<u>TTTGATTCTCCAATT</u>GACAG	CCGCCACATCCGCAT
55.7	+	CCACCCGGGTAATA**a**AGTCAG<u>TTTCCTT</u>GGA<u>TCCTCCATT</u>ACAG	GTCAGTTCATCGGGG
52.2	+	TTTCATCATCATACC**a**T<u>CTTCATCTTCGTCTTCATCCTCTT</u>CAG	CAAAAGACCAGAACA
53.6	−	GCAACTCAGGACGCCCTCTGACTA**a**<u>TCTCTTTCAATTCTT</u>GCAG	CATGCTACCAACTGG
54.7	−	TCTAATCA**a**TA<u>CCCCTTCCTCCCACTT</u>CAA<u>TTTGTGTCTT</u>GCAG	AACAGGAGGATGACT
52.3	+	TTTCATCATCATACC**a**TCTTCA<u>TCTTCGTCTTCATCCTCTT</u>CAG	CAAAAGACCAGAACA
51.4	+	CCACCAGTTTGTACTAACCA**a**GTGTGTATA<u>TTTTCTATTC</u>ACAG	GTTCTGTTGCGTACT
59.1	+	GACTTTACACTTAAGCTAACTCAAAGCTA**a**<u>TTTGTTTCTT</u>ACAG	TATGCGATGATCTGG
54.6	+	TTGCGTTGCCTTTAACTA**a**TTCTG<u>CCTGTTTTTTCTATGTTTT</u>AG	TGACATTTGATGTGT
56.3	+	ATGTATAACATTTTGCATTCACACTA**a**<u>TTTGTTTTATTTTTC</u>AG	ATTTAGGTGAACTCA
54.2	+	TCCCAAGTTTATGGTTATTCTA**a**TTA<u>TCTGCTTAATCCTTTT</u>AG	AACTGGCTGTCTTCG
46.3	+	GTTATAAACTAAAGTA**a**CGACGTA<u>TTCCTTTGTCCCACC</u>AACAG	GATGTGACGGTCTGC
57.4	+	AACAAGTTACTA**a**ATTCA<u>TTCTTCTGTATTC</u>GGG<u>TTTCTTT</u>CAG	AATTCGATTACGGAC
58.6	+	CCAACACTTTCGACCACATCTA**a**TA<u>CCATCTTTCTTTATT</u>GCAG	TTGTTCTGGCCAACG
54.2	+	ATTAGGTAACCCATACTA**a**<u>CCTTTTACATTTCCACTCTTTTT</u>AG	AAAAACAATTCGAGT
52.0	+	ACTTTGTGTGTATATATCATTA**a**CA<u>CTTATGCTTTCCTTT</u>GCAG	GTACCTCACATTTGG

Fig. 4. Alignment of 3′-intron sequences upstream of the specific exons 5 of all identified alternatively spliced *mod(mdg4)* mRNAs. All sequences contain the conserved dinucleotide AG at the 3′-boundary. Pyrimidine-rich tracts are underlined and the putative branch point is indicated by a small bold letter. Specific exons encoded by the same DNA strand as exons 1–4 are indicated by (+) and those encoded by the cDNA strand by (−).

of any GAL4-driver element (Fig. 6C). Independent of the insertion site of the transgene and its orientation relative to the UAS sequence, trans-splicing of the tagged specific exon *mod-(mdg4)-55.1* could be demonstrated. For PCR primers, again the forward primer E4-F and the primer 55.1-tag1-back have been used. However, the level of expression and/or the efficiency of trans-splicing is variable in different transformants. This could be because of chromosomal position effects, depending on the insertion site of the transgene. We conclude from these results that independent mRNAs containing the common exons 1–4 on one hand and mRNAs containing the specific exon *mod(mdg4)-55.1* on the other hand are produced endogenously.
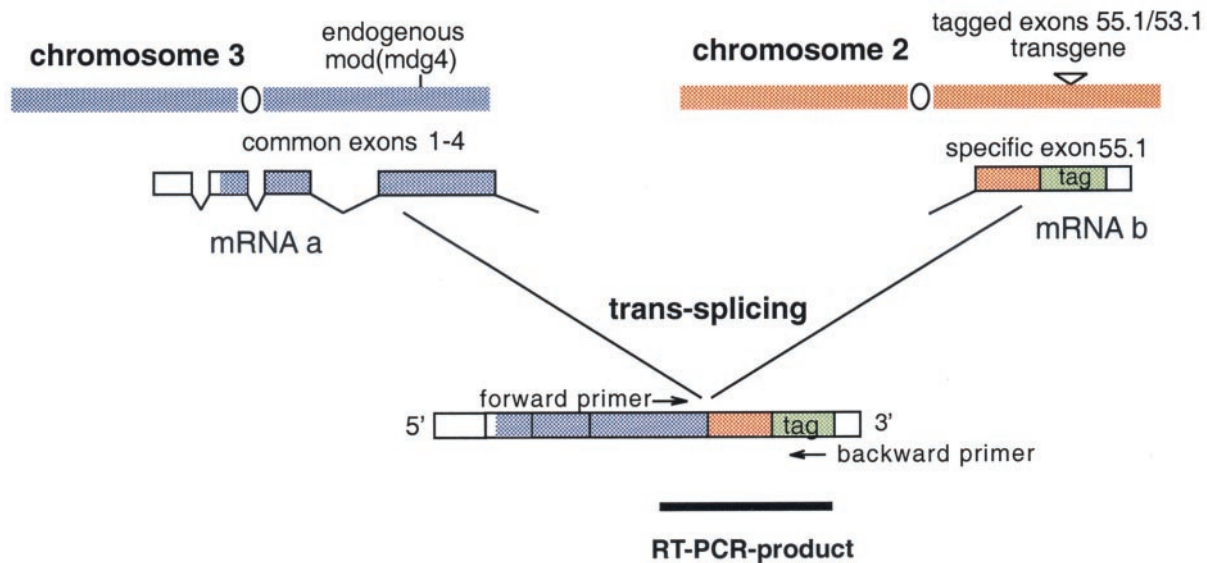

Fig. 5. Transgenic assay to demonstrate the existence of trans-splicing at *mod(mdg4)*. Schematic representation of the trans-splicing assay in case of a second chromosomal transgene. Common exons of *mod(mdg4)* are expressed from the endogenous locus producing mRNA a. The sequence-tagged specific exon *mod(mdg4)-55.1* is expressed under control of GAL4 binding to the UAS upstream of the transgene, producing mRNA b. The mature trans-spliced mRNA is detected by RT-PCR with the indicated pair of primers.
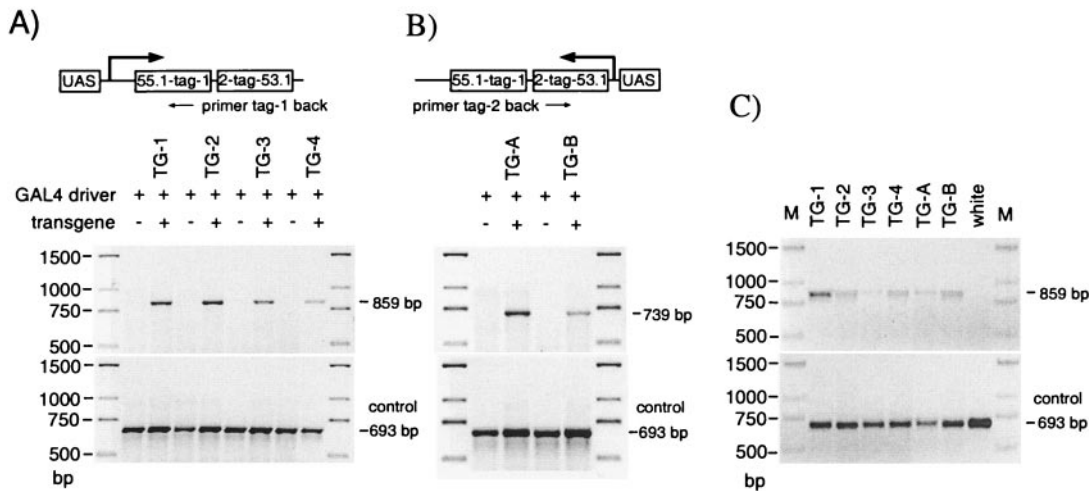
Fig. 6. RT-PCR demonstrating trans-splicing of endogenous *mod(mdg4)* common exons 1–4 and tagged-specific *mod(mdg4)* exons expressed from a transgene. (*A*) Expression of the tagged exon *mod(mdg4)-55.1* is driven by crossing heterozygous transgenic lines to homozygous *nanos*-GAL4-producing driver lines. The amplicon of the expected size is only obtained in case of trans-splicing of endogenous exon 4 and tagged exon 5 expressed from the transgene. The primer E4-F and tag1-back have been used for PCR. As a control for RT-PCR a backward primer complementary to exon 5 upstream of the sequence tag was used in combination with E4-F (*Lower*). (*B*) RT-PCR results obtained with transgenic lines expressing the tagged-specific exon *mod(mdg4)-53.1* under control of the nanos GAL4-driver line. The tag2-back primer has been used in combination with primer E4-F. Only in case of correct trans-splicing of the two independent transcripts, the expected amplicon is obtained. No amplicon is obtained in absence of the transgene. For control of RT-PCR the same primer pair as in *A* was used (*Lower*). (*C*) Demonstration of a sequence element with putative promoter activity upstream of specific exon 5 of *mod(mdg4)-55.1* contained in the transgene. RT was performed with total RNA from females containing the transgenes TG-1 to 4 (pUAST-55.1-tag1-tag2–53.1) and TG-A and -B (pUAST-53.1-tag2-tag1–55.1) in the absence of a GAL4-driver element. For PCR the forward primer E4-F and the back primer tag1-back were used. In all transgenic lines, a PCR fragment of the expected size was obtained. Quantitative differences among the transgenic lines could be because of position effects. No fragment was amplified with total RNA isolated from the strain *w^1118*. As a control for RT-PCR, the backward primer deduced from the specific exon *mod(mdg4)-55.1* upstream of the sequence tag was used in combination with the E4-F primer.

Trans-Splicing of the Common Exons 1–4 Is Independent of Their Proximity to the 3′-Specific Exons. If trans-splicing is a general property of *mod(mdg4)*, we would expect that separation of common exons 1–4 and 3′-specific exons via transgene insertion should result in the generation of mRNAs containing endogenous specific exons. To test this we made use of a second chromosomal transgene containing the 7.5 kb-*Bam*HI genomic fragment (at position 0–7.5 according to the physical map, Fig. 2). This transgene was shown to partially rescue the recessive lethality of the *mod(mdg4)^neo129* mutation. This mutation is caused by insertion of the pUChsneory⁺ transposon within the third intron (11). Sequence analysis revealed a DNA polymorphism at nucleotide position 555 in exon 4 (C in transgene, T in mutant chromosome), which allows detection of mRNAs orig-
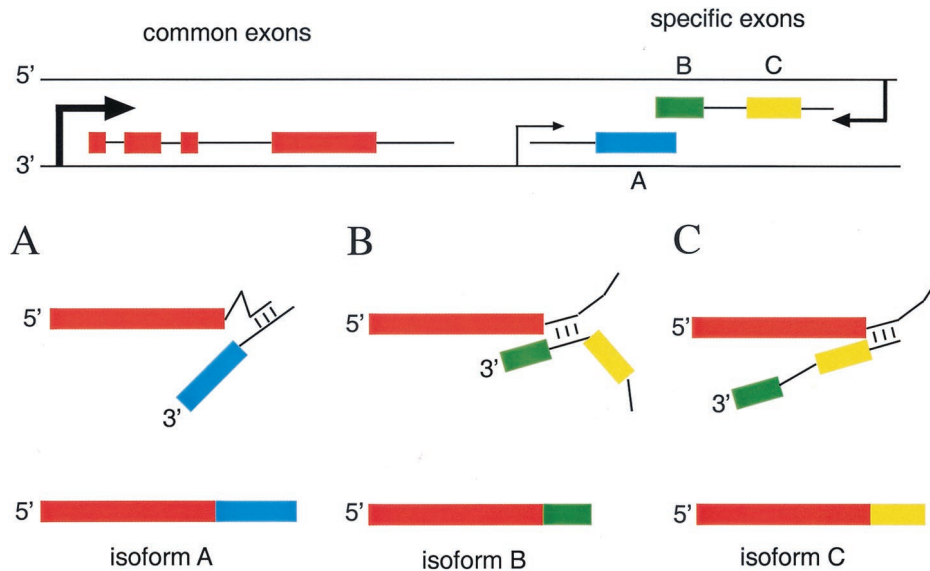

Fig. 7. The model demonstrating different mRNA trans-splicing events at the *mod(mdg4)* locus. Common exons 1–4 located at the 5′ end of the locus are supposed to be strongly transcribed as a unique transcription unit. Additional transcription units containing one (*A*) two (*B* and *C*) or even more (not shown) isoform-specific exons are initiated at the same coding strand (*A*) or at the antiparallel DNA strand (*B*/*C*). In case of mRNAs containing two or more specific exons, different 5′ splice sites can be used, resulting in the generation of different mRNA variants (*B* vs. *C*). The relative amount of single *mod(mdg4)* protein isoforms could be regulated by the rate of transcription initiation, by the efficiency of trans-splicing, and by the stability of the resulting mRNA.

inated from the transgene. Trans-splicing of transgene encoded common exons 1–4 and endogenous 3′-specific exons was assayed by RT-PCR with total RNA isolated from females containing the transgene and which are homozygous for the *mod(mdg4)^neo129* mutation. To prove that the specific exon *mod(mdg4)-55.1* is correctly spliced to exon 4, PCR was performed with primer E4-F and primer 55.1-RT-B. Sequencing of four cloned PCR fragments obtained in two independent RT-PCR experiments indicated that one clone contains the transgene-encoded sequence polymorphism. This result demonstrates correct trans-splicing of common exon 4 to the endogenous 3′ exon of isoform *mod(mdg4)-55.1* independent of their chromosomal location.

Discussion

In this study, we provide experimental proof for the occurrence of mRNA trans-splicing in *Drosophila*. The mod(mdg4) locus represents an unusual type of gene structure. Both DNA strands within the locus are used to encode a large number of protein isoforms. Our transgenic approach clearly demonstrates that both the colinearly located specific exons (demonstrated for exon *mod(mdg4)-55.1*) and those encoded by the antiparallel DNA strand (shown for exon *mod(mdg4)-53.1*) are substrates for trans-splicing. This result also suggests that all other *mod(mdg4)* isoforms might be generated by trans-splicing, implying the initiation of independent pre-mRNAs at several promoter elements within the mod(mdg4) complex. We indeed found multiple TATA-box-containing elements throughout the locus. One of these is located upstream of the *mod(mdg4)-55.1* isoform and is contained in the transgene. Expression of the transgene independent of the inducible promoter element in six independent transgenic lines indicates a putative promoter function. However, further experiments should demonstrate the existence of multiple promoters at the *mod(mdg4)* locus. Moreover, our results demonstrate that trans-splicing occurs within *mod(mdg4)*,

independent of the chromosomal context of the common exons 1–4 and the 3′-specific exons. This also raises the question about the special requirements for initiating trans-splicing at *mod(mdg4)*. Further experiments should clarify whether RNA recognition or nuclear compartmentalization or both plays a role in the initiation of trans-splicing.

Our data suggest that trans-splicing is a general property of the *mod(mdg4)* locus. Three possible types of trans-splicing events are shown by the model presented in Fig. 7. The transcript containing common exons 1–4 (shown in red) is produced in large quantities and contains putative interaction sites with upstream regions of pre-mRNAs containing the specific mod-(mdg4) exon(s). The specific exons are transcribed as monoexonic, di-exonic, or polyexonic mRNAs from both cDNA strands. Depending on the site of trans-splicing, three different protein isoforms (A, B, and C) are produced. The expression of the alternatively spliced *mod(mdg4)* isoforms could be regulated at several levels: (*i*) differential spatial and temporal expression of pre-mRNAs containing one (or groups of) alternatively spliced specific exons; (*ii*) differences in selectivity and efficiency of trans-splicing to generate different quantities of mature mRNAs, and (*iii*) variable stability of the isoform specific transcripts. As a result, >20 different *Mod(mdg4)* protein isoforms are produced that all contain a common region of 402 aa, including the N-terminal BTB domain implicated in dimerization/oligomerization (18, 19) and variable C termini with the conserved C_2H_2 motif (11). The variable C termini are implicated to specify the function of individual isoforms in different processes like chromatin insulator function (9, 20), programmed cell death (10), or modification of gene silencing (8).

Our results demonstrate the existence of trans-splicing in *Drosophila* and provide a new basis for a genetic and molecular analysis of key elements involved in this process.

1. Murphy, W. J., Watkins, K. P. & Agabian, N. (1986) *Cell* **47,** 517–527.
2. Blumenthal, T. (1995) *Trends Genet.* **11,** 132–136.
3. Garcia-Blanco, M. A., Puttaraju, M., Mansfield, S. G. & Mitchell, L. G. (2000) *Gene Ther. Regul.* **1,** 141–163.
4. Eul, J., Graessmann, M. & Graessmann, A. (1995) *EMBO J.* **14,** 3226–3235.
5. Li, B.-L., Li, X.-L., Duan, Z.-J., Lee, O., Lin, S., Ma, Z.-M., Chang, C. C. Y., Yang, X.-Y., Park, J. P., Mohandas, T. K., *et al*. (1999) *J. Biol. Chem.* **274,** 11060–11071.
6. Caudevilla, C., Serra, D., Miliar, A., Codony, C., Asins, G., Bach, M. & Hegardt, F. G. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12185–12190.
7. Labrador, M., Mongelard, F., Plata-Rengifo, P., Baxter, E. M., Corces, V. G. & Gerasimova, T. I. (2001) *Nature (London)* **409,** 1000.
8. Dorn, R., Krauss, V., Reuter, G. & Saumweber, H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11376–11380.
9. Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O. & Corces, V. G. (1995) *Cell* **82,** 587–597.
10. Harvey, A. J., Bidwai, A. P. & Miller, L. K. (1997) *Mol. Cell. Biol.* **5,** 2835–2843.
11. Büchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G. & Dorn, R. (2000) *Genetics* **155,** 141–157.
12. Read, D., Butte, M. J., Dernburg, A. F., Frasch, M. & Kornberg, T. B. (2000) *Nucleic Acids Res.* **28,** 3864–3870.
13. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
14. Brand, A. & Rerrimon, N. (1993) *Development (Cambridge, U.K.)* **118,** 401–415.
15. Rubin, G. M. & Spradling, A. C. (1982) *Science* **218,** 348–353.
16. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
17. Zollmann, S., Godt, D., Prive, G. G., Couderc, J. L. & Laski, F. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10717–10721.
18. Ahmad, K. F., Engel, C. K. & Prive, G. G. (1998) *Proc. Natl Acad. Sci. USA* **95,** 12123–12128.
19. Bardwell, V. J. & Treisman, R. (1994) *Genes Dev.* **8,** 1664–1677.
20. Bell, A. C., West, A. G. & Felsenfeld, G. (2001) *Science* **291,** 447–450.

GENETICS