Trans-splicing of the mod(mdg4) complex locus is conserved between the distantly related species *Drosophila melanogaster* and *Drosophila virilis*

Manuela Gabler*, Michael Volkmar*, Susan Weinlich*, Andreas Herbst*, Philine Doebberthien*, Stefanie Sklarss*, Laura Fanti†, Sergio Pimpinelli†, Horst Kress††, Gunter Reuter* and Rainer Dorn*

* Institute of Genetics, Martin Luther University, D-06120 Halle, Germany
† Department of Genetics and Molecular Biology, University La Sapienza, I-00185 Rome, Italy
†† Institute for Biology-Genetics, Free University of Berlin, D-14195 Berlin, Germany
Running head: evolutionary conservation of trans-splicing

Keywords: *Drosophila virilis*, *mod(mdg4)*, trans-splicing, evolutionary conservation

**corresponding author:** Rainer Dorn

mailing adress: Institute of Genetics, Martin Luther University, Weinbergweg 10, D-06120 Halle, Germany
Phone: 49-(0)-345-55-26323
Fax: 49-(0)-345-55-27294

E-mail: dorn@genetik.uni-halle.de
ABSTRACT

The modifier of mdg4, mod(mdg4), locus in Drosophila melanogaster represents a new type of complex genes, where functional diversity is resolved by mRNA trans-splicing. A protein family of more than twenty transcriptional regulators, which are supposed to be involved in higher order chromatin structure, is encoded by both DNA strands of this locus. Mutations in mod(mdg4) have been identified independently in a number of genetic screens, like position effect variegation, modulation of chromatin insulators, apoptosis, pathfinding of nerve cells and chromosome pairing, indicating pleiotropic effects.

The unusual gene structure and mRNA trans-splicing are evolutionary conserved in the distantly related species Drosophila virilis. Chimeric mod(mdg4) transcripts encoded from nonhomologous chromosomes comprising the splice donor from D. virilis and the acceptor from D. melanogaster are produced in transgenic flies. We demonstrate that significant amount of protein can be produced from these chimeric mRNAs. The evolutionary and functional conservation of mod(mdg4) and mRNA trans-splicing in both Drosophila species is furthermore demonstrated by the ability of D. virilis mod(mdg4) transgenes to rescue recessive lethality of mod(mdg4) mutant alleles in D. melanogaster.
INTRODUCTION

The majority of genes in higher eukaryotes represents monocistronic units where noncoding intron regions interrupt the protein-coding exon sequences. The resulting mature mRNA encodes usually a unique polypeptide. Recent advances in genome analysis of several model organisms and the molecular characterization of a large number of genes revealed, that alternative pre-mRNA splicing is one of the main mechanisms generating a highly expanded proteome diversity. Thus, protein families with slightly different isoforms or even proteins with unrelated functions can be produced from single or multiple promoter elements within one gene (for review see GRAVELEY 2001; MANIATIS and TASIC 2002). Regulatory integration of different transcriptional units is found in gene complexes like Hox genes, hemoglobin genes or immunoglobin genes. This organisation reflects clustering of genes with related functions. With mod(mdg4) a new type of functional clustering has been discovered in Drosophila. This complex locus encodes more than two dozens isoforms generated by mRNA trans-splicing (BÜCHNER et al. 2000; DORN et al. 2001; KRAUSS and DORN 2004). Protein isoforms produced by mod(mdg4) contain a common 402 amino acid N-terminal region encoded by the four 5’ exons but differ in their C-terminal region encoded by alternative 3’ exons. This kind of trans-splicing clearly differs from SL (splice leader) trans-splicing predominating in Ceanorhabditis and Trypanosomes where polycistronic transcripts are resolved by addition of noncoding leader sequences (BLUMENTHAL 1998). Mutational dissection and differential binding of Mod(mdg4) isoforms on polytene chromosomes suggest that the variable C-terminal regions encoded by any of the alternative 3’ exons determine functional specificity. Specific Mod(mdg4) isoforms are supposed to be involved in control of heterochromatic gene silencing, regulation of homeotic genes, function of chromatin insulators, nerve cell pathfinding, induction of apoptosis and control of meiotic processes (DORN et al. 1993; GERASIMOVA et al. 1995;
HARVEY et al. 1997; GORZYCA et al. 1999; BÜCHNER et al. 2000). Genomic structure and transgene analysis demonstrate the specific functional organisation of the complex mod(mdg4) locus (DORN et al. 2001). Mature mod(mdg4) transcripts are generated by a trans-splicing mechanism combining one primary transcript comprising the common four 5’ exons with another transcription unit contributing one of the alternative 3’ exons. A comparably complex gene structure was also described for a number of other genes in Drosophila, including Broad, tramtrack, GAGA-factor/Trl and lola, all of which encode numerous protein isoforms with alternative C-termini (DiBELLO et al. 1991; READ and MANLEY 1993; SOELLER et al. 1993; MADDEN et al. 1999). Interestingly, besides mod(mdg4), mRNA trans-splicing was recently reported for the lola locus (HORIUCHI et al. 2003). Another unique characteristic of these genes is that they all encode BTB/POZ domain proteins, which frequently contain Cys₂His₂ zinc finger motifs within the variable C-terminal region. Therefore, mod(mdg4) appears to represent a prototype of a new class of complex loci where functional variety is produced by a trans-splicing mechanism combining independent transcription units in a regulated manner.

Towards a better understanding of the mod(mdg4) complex structure and the underlying conserved principles involved in mRNA trans-splicing we analyzed the Drosophila virilis orthologue of mod(mdg4). Molecular analysis of D. virilis mod(mdg4) revealed strong conservation of gene structure and the encoded protein isoforms. Functional conservation is suggested by similar protein distribution of orthologue isoforms on polytene chromosomes in both species and by mutant rescue in transgenic Drosophila melanogaster lines after expression of D. virilis mod(mdg4) protein isoforms. Moreover, we show that mRNA trans-splicing is conserved in both species allowing the generation of chimeric D. melanogaster/D. virilis mod(mdg4) transcripts in vivo. Quantitative RT-PCR experiments
reveal that mod(mdg4)-67.2 chimeric transcripts consisting of a donor encoded by a third chromosomal transgene and the acceptor encoded by the third chromosomal endogenous locus represents about 12% with respect to the endogenous D. melanogaster mod(mdg4)-67.2 transcript. The corresponding chimeric protein can be detected at nearly wildtype level on polytene chromosomes of transgenic mod(mdg4) homozygous larvae. Our data provide new insight into functional conservation of a new type of complex loci where functional complexity is resolved by mRNA trans-splicing.
MATERIALS AND METHODS

Library screen:
Screening of a D. virilis genomic library (LANIO et al. 1994) was performed as described in SAMBROOK et al. (1989). Washes were performed twice in 2x sodium sodium citrate (SSC), 0.1% sodium dodecyl sulphate (SDS) for 10 min. Two cDNA clones corresponding to isoforms mod(mdg4)-67.2 and mod(mdg4)-58.8 were used as radiolabeled hybridization probes. Three recombinant genomic λ clones were isolated and the genomic insert of the representative clone λDv2-1 was sequenced and used in all other experiments. The overlapping recombinant λ clone Dv3-2 was isolated in a second library screen with the help of a probe deduced from λDv2-1.
The GenBank/EMBL accession number of the D. virilis mod(mdg4) genomic sequence is AJ586737.

Construction of transgenic lines:
The 11.5kb genomic insert of clone λDv2-1 was cloned into the unique NotI site of the P-transformation vector pW8 (KLEMENTZ et al. 1986). After transformation three independent transgenic lines (1., 2. and 3. chromosome) have been obtained. The transgene was designated as P(w+ Dv mod(mdg4) 11.5kb NotI) and arabic numbers indicate the chromosomal location of the transgene. The genomic 6.8kb NotI-XbaI fragment containing exclusively the common exons 1 to 4 was cloned into pW8 and one second chromosomal transgenic semilethal line was established. Transformation of D. melanogaster was performed as described by RUBIN and SPRADLING (1982).
The chromosomes 3-P(w+ Dv mod(mdg4) 11.5kb NotI) mod(mdg4)neo129 and 3-P(w+ Dv mod(mdg4) 11.5kb NotI) mod(mdg4)02 were obtained by recombination. The presence of both, the transgene and the mod(mdg4) mutations was tested with specific PCR primer pairs. Recombinant chromosomes, which lost the w+ marked transgene did not complement the original mod(mdg4) mutations.

**Drosophila strains and crosses:**

*mod(mdg4)* mutant alleles are described in BÜCHNER et al. (2000). The allele *mod(mdg4)07/351* was obtained from M. Frasch. Other strains are described in LINDSLEY and ZIMM (1992).

All crosses were performed at 25°C. For complementation analysis *mod(mdg4)* mutant strains containing the *wm* chromosome and the *TM3, Sb Ser* balancer have been used. One copy of the *D. virilis mod(mdg4)* transgene was inherited maternally and its presence in offspring flies was monitored by the *w+* marker gene. The percentage of homozygous/transheterozygous *mod(mdg4)* mutant flies was calculated as a percentage of the expected number of these flies (only flies containing the transgene were included in this calculation) in the appropriate crosses. All mutant alleles do not complement in absence of the transgenes under the conditions used. This is confirmed by the absence of homozygous/transheterozygous *mod(mdg4)* mutant flies without the *D. virilis* transgene in all crosses.

**RNA isolation and RT-PCR:**

Isolation of poly(A)+ RNA was performed as described in DORN et al. (2001). Primers used in RT-PCR experiments are: ex4-virF: 5’ CGAGCACCAGCCAACGTAATTGATC 3’, 64.2-B-RT:
5' CAA/gCTTGAGC/eTCCTTGCCG/aTC 3' (different nucleotide positions in D. virilis are indicated by small letters), 51.4-B-RT: 5' CAAGACCAATAAGTTTCAATCCCG 3', 56.3-B-RT:
5' ACATCGCCCGCTCTGGGTC 3'
The new isoform mod(mdg4)-53.5 was identified with primers ex4-melF: 5'
CGCAAATGTTATGGGACCCTCTC 3' and 53.5-RT-Bmel: 5'
CGGCTTGTGATTGTGAAATCC 3' in D melanogaster and primers ex4-virF: 5'
CGAGCACCACGCAACGTAATTGAC 3' and 53.5-RT-Bvir: 5'
GTAATCCTGAGCTCTGTGAGC 3' in D. virilis. Total RNA was isolated with TRIzol (Invitrogen) and poly(A)⁺ RNA was obtained with the mRNA Purification Kit (Amersham).
For reverse transcription (RT) 1µg RNA was incubated with random hexamer primer and Moloney murine leukemia virus reverse transcriptase (Promega) according to the manufacturer’s protocol. PCR reaction mixture contained 167µM dNTPs, 1.67mM MgCl₂, 267nM of each primer and 1.5 units Taq DNA polymerase. Conditions for PCR were 95°C for 5min, 95°C for 40sec, 55°C for 40sec and 72°C for 40sec (35 cycles).

**Real-time quantitative RT-PCR:**
Total RNA was extracted from adult female flies of strains w¹¹¹⁸, w¹¹¹⁸ 3-P(w⁺ Dv mod(mdg4) 11.5kb NotI)/3-P(w⁺ Dv mod(mdg4) 11.5kb NotI) and w¹¹¹⁸ 3-P(w⁺ Dv mod(mdg4) 6.8kb NotI-XbaI)/3-P(w⁺ Dv mod(mdg4) 6.8kb NotI-XbaI) using the TRIzol reagent (Invitrogen). PCR primers were selected intron-spanning using the GeneRunner software (Hastings Software, Inc.) The primer sequences are as follows: rp49-fwd: 5'
TGTCCTTCCAGCTTCAAGATGACCAC 3'; rp49-rev: 5'
CTTGGGGCTTGCGCCATTTGTG 3'; mel-com-fwd: 5'
TTCTTCCGCAAGATGTTCACATAGATG 3'; mel-com-rev: 5'
TGAATTGGATGAGGTCCTTCAGCG 3’; vir-com-fwd: 5’
CGCACCGTTTGGTGTCTGTCTGC 3’; vir-com-rev: 5’
CTTATCAGGTCTTCAATGCGGAATGGC 3’; mel-spec-fwd: 5’
CAAATACGACCGGTGCAGGAGTGAC 3’; vir-spec-fwd: 5’
ACACAACACCACCAGCTCAAGGC 3’; mel-64.2-rev: 5’
TGGCTGCAAATGAAACTGATCTCCG 3’; vir-64.2-rev: 5’
GTTGTCGCCCCATTGCTTGGGTC 3’; mel-67.2-rev: 5’
TTTCGGTGCTGCCGTTACGTG 3’. Specificity of the primers was confirmed by melting curve analysis, agarose gel electrophoresis and sequencing of the PCR products.
The real-time qRT-PCR was performed with Quantitect SYBR Green RT-PCR Kit (QIAGEN) on an iCycler Thermocycler (Bio-Rad) in triplicate for each sample and analyzed with the iCycler iQ software (Bio-Rad). For data standardization, the absolute expression level of each mRNA was determined and set in proportion to D. melanogaster rp49, which is strongly expressed in females (TAMATE et al. 1990).

**Immunological analysis:**
Staining of polytene chromosomes was performed as described in BÜCHNER et al. (2000). Immunostaining was performed with anti-Mod(mdg4)-58.0<sup>BTB-534</sup> (1:1000 dilution) which recognizes all Mod(mdg4) isoforms and anti-Mod(mdg4)-58.0<sup>403-534</sup> (1:100 dilution) which detects exclusively the isoform Mod(mdg4)-58.0. The antibody anti-Mod(mdg4)-67.2<sup>403-610</sup> was generously provided by P. Geyer (Iowa) and used in a 1:1000 dilution on polytene chromosomes. The presence of the D. virilis transgene was probed by PCR with a D. virilis specific primer pair (primer D.vir 7211F: 5’ TGATGTAAGTTGGGTTCCATTGCG 3’, primer D.vir 64.2B: 5’ GGATCCATGCAGCTTGAAGCTTGTGCAG 3’), using DNA isolated from the corresponding carcasses as template.
Immunofluorescence and sequential in situ hybridization with salivary gland chromosomes was performed as described in PIMPINELLI et al. (2000).

**Western analysis:**

Salivary glands of w1118, mod(mdg4)02/mod(mdg4)02, 2-P(w+ Dv mod(mdg4) 11.5kb NotI)/+; mod(mdg4)02/mod(mdg4)02 and P-(w+ Dv mod(mdg4) 6.8kb NotI-XbaI)/+; mod(mdg4)02/mod(mdg4)02 were dissected from third instar larvae in IP-Buffer (20mM Tris/HCl pH8.0; 150mM NaCl; 10mM EDTA; 1mM EGTA; 2mM Na2VO4). Equal protein amounts were loaded and western blots were probed with polyclonal anti-Mod(mdg4)-67.2403-610 antibody (1:1000) and monoclonal anti-Tubulin antibody (Sigma; 1:20,000).
RESULTS

The genomic structure of the mod(mdg4) locus in D. melanogaster and D. virilis is highly conserved:

Recently, we have described the genomic structure of the mod(mdg4) locus in D. melanogaster. The structural organisation of this complex locus is unusual because both DNA strands are used as coding strands and independent transcripts are combined via mRNA trans-splicing (DORN et al. 2001; LABRADOR et al. 2001). To prove, if mod(mdg4) is conserved in distantly related Drosophila species we tested the polyclonal antiserum anti-Mod(mdg4)-58.0\textsuperscript{BTB-534} for crossreactivity in D. virilis. This antibody stains a number of several hundred bands in salivary gland polytene chromosomes of D. melanogaster (BÜCHNER et al. 2000). A similar effect was obtained if polytene chromosomes of D. virilis were stained with the same antiserum (data not shown). In order to prove the conservation of individual Mod(mdg4) isoforms and their chromosomal binding we stained D. virilis polytene chromosomes with the specific antibody anti-Mod(mdg4)-58.0\textsuperscript{403-534} detecting exclusively the isoform Mod(mdg4)-58.0. This isoform is significantly less abundant and restricted to about 25 reproducibly stained sites on D. melanogaster polytene chromosomes (BÜCHNER et al. 2000). A comparably low number of stained sites is detected in polytene chromosomes of D. virilis (Figure 1A). However, the decreased number of stained sites in D. virilis may be due to a differential distribution and/or to lower affinity of the antibody in D. virilis. Therefore we tested two individual binding sites, which are known to comprise related genes in the two species. Subdivision 33A in D. virilis harbours the gene vE74, the orthologue of the D. melanogaster E74A gene (JONES and DALTON 1991) corresponding to the anti-Mod(mdg4)-58.0\textsuperscript{403-534} binding site 74EF. A second corresponding site is subdivision 58E in D. virilis, which carries one of the three
clusters of 5S rRNA genes (KRESS 2001). In *D. melanogaster* the 5S RNA genes are clustered in subdivision 56E (WIMBER and STEFFENSON 1970). Both are strong binding sites for isoform Mod(mdg4)-58.0. Another example of common binding sites of Mod(mdg4) is the *Bithorax-Complex (BX-C)*. The observed homeotic transformation of the mutant allele *mod(mdg4)neo129* indicated an involvement of *mod(mdg4)* in the transcriptional regulation of the *BX-C* (DORN et al. 1993). To prove, if Mod(mdg4) binds to the *BX-C* of both species we performed sequential staining with anti-Mod(mdg4)-58.0<sup>BTB-534</sup> and a probe of *D. melanogaster* *Ubx* DNA on polytene chromosomes. In fact anti-Mod(mdg4)-58.0<sup>BTB-534</sup> binds close to the *Ubx* probe in both species (Figure 1B), indicating its binding to homologous sites. These results clearly point to the structural and functional conservation of the *mod(mdg4)* locus in *D. virilis*.

For isolating the *D. virilis mod(mdg4)* locus we selected two cDNA clones corresponding to isoforms *mod(mdg4)-67.2* and *mod(mdg4)-58.8* as hybridization probes and screened a genomic *D. virilis* λlibrary. Three independent recombinant λclones were isolated and restricted, among other endonucleases, with *SalI*. All recombinant clones contained a 0.5kb *SalI* restriction fragment, which is also present in the common exon four of *D. melanogaster mod(mdg4)*. Sequence analysis of the cloned fragment revealed a significant sequence conservation within this coding region. Subsequently, we sequenced the genomic insert of one representative λclone. Comparison with the corresponding sequence of *D. melanogaster mod(mdg4)* revealed a strong conservation of both, genomic structure and Mod(mdg4) isoforms (Figure 2). Most importantly, also in *D. virilis* the orthologous isoforms Mod(mdg4)-53.1, Mod(mdg4)-62.3, Mod(mdg4)-55.6, Mod(mdg4)-53.6, Mod(mdg4)-54.7, Mod(mdg4)-59.0 and Mod(mdg4)-67.2 are encoded by the opposite DNA strand. These data strongly imply that mRNA trans-splicing is also conserved in both species. To verify, if the
putative coding regions are transcribed in *D. virilis*, we performed RT-PCR experiments for six selected *mod(mdg4)* isoforms, *mod(mdg4)-64.2*, -60.1, -55.1, -53.1, -58.0 and -67.2, respectively. The forward primer corresponds to the putative common exon 4 and the backward primers are isoform specific primers hybridizing downstream of the putative open reading frames (ORF). The resulting fragments were sequenced and the predicted ORFs and the alternative splice sites verified. The exon/intron structure of the common 5’ region has been determined by RT-PCR with a forward primer deduced from the first coding exon and a backward primer deduced from exon 4.

**Comparison of the *D. melanogaster* and *D. virilis* Mod(mdg4) protein isoforms:**

The multiple *D. melanogaster* Mod(mdg4) protein isoforms contain a common N-terminal region of 402 amino acids encoded by exons 2 to 4 and variable, isoform-specific, C-termini encoded by alternatively spliced exon 5 or exons 5 and 6 (BÜCHNER *et al.* 2000). Comparison of the common region (Figure 3) reveals a strong conservation (82% identity). The N-terminal BTB/POZ domain is identical in both species. Within the remaining common protein part (amino acids 121 to 413) only a small number of amino acid replacements (in most cases conservative ones) are present. The difference in size (413 versus 402 amino acids) is mainly attributed to additional glutamine residues in *D. virilis*. Next we compared the specific C-termini of Mod(mdg4) isoforms. Figure 3 also shows the alignment of specific C-terminal regions of five Mod(mdg4) isoforms. Despite the strongly conserved amino acid positions of the FLYWCH consensus sequence most of the amino acid positions within this motif are also conserved between orthologous isoforms. In case of isoform Mod(mdg4)-67.2 the identity is 85% and isoforms Mod(mdg4)-64.2 and
Mod(mdg4)-60.1 show an identity of 90% within the FLYWCH motif. Based on this observation we theorize that these amino acids mainly contribute to specific functions of the different Mod(mdg4) isoforms. Outside the FLYWCH domain the sequence of the orthologous isoforms is less but still conserved. The isoforms Mod(mdg4)-55.1 and Mod(mdg4)-58.0 do not contain the FLYWCH consensus sequence. However, also in these cases the specific C-termini are significantly conserved (45% and 51% identity, respectively). These results support our view that the different orthologous Mod(mdg4) protein isoforms have individual functions which are conserved in both Drosophila species. A comprehensive evolutionary analysis of all Mod(mdg4) protein isoforms identified in *D. virilis* and other Diptera species is presented in KRAUSS and DORN (2004).

*mod(mdg4)* is expressed in all stages of development and is maternally provided in both Drosophila species:

To compare the developmental expression pattern of *mod(mdg4)* in both Drosophila species we performed Northern blot analyses. In a previous analysis with poly(A)+ RNA we detected abundant transcripts at 2.0kb and 2.3kb in all stages of development (BÜCHNER et al. 2000). As expected, *D. virilis* *mod(mdg4)* shows a very similar expression pattern, characterized by abundant transcripts during embryonic development and in females (data not shown). To prove, if maternal Mod(mdg4) proteins are present in preblastoderm embryonic stages we performed antibody staining with anti-Mod(mdg4)-58.0<sup>BTB-534</sup> which detects all Mod(mdg4) protein isoforms. In both species Mod(mdg4) proteins could be detected during early cleavage cycles (data not shown).

*Drosophila virilis* *mod(mdg4)* transgenes rescue mutant phenotypes in *D. melanogaster*:
The strong structural conservation of mod(mdg4) in *D. virilis* prompted us to test its functional conservation. Therefore, we established transgenic lines containing the genomic insert of the recombinant phage λDv2-1 which encodes the *D. virilis* mod(mdg4) common exons 5 specific exons from isoforms mod(mdg4)-64.2, mod(mdg4)-60.1, mod(mdg4)-53.5, mod(mdg4)-55.1, mod(mdg4)-53.1 and partial specific exon from mod(mdg4)-62.3 (cf. Figure 2). This 11.5kb genomic NotI fragment is expected to produce these 5 *D. virilis* Mod(mdg4) protein isoforms, but not the Mod(mdg4)-62.3 because both, the trans-splice site and a putative promoter driving its expression are not included. This prediction was confirmed by RT-PCR experiments with two independent transgenic lines. To prove the capability of the *D. virilis* P(w+ Dv mod(mdg4) 11.5kb NotI) transgene to rescue the recessive lethality of mod(mdg4) mutant alleles, strains containing the *D. virilis* transgene and one of the two recessive lethal alleles, mod(mdg4)\(^{02}\) and mod(mdg4)\(^{neo129}\), have been constructed. These strains have been used in complementation crosses with mutant alleles mod(mdg4)\(^{neo129}\), mod(mdg4)\(^{R32}\), mod(mdg4)\(^{02}\) and mod(mdg4)\(^{07}\). All alleles represent mutations within the mod(mdg4) common region and do not complement each other in the absence of the transgene (BÜCHNER et al. 2000; this work). The P(w+ Dv mod(mdg4) 11.5kb NotI) transgene is able to rescue almost all trans combinations of mod(mdg4) mutant alleles used, independently whether the transgene is located on the second or the third chromosome (Table 1). However, mod(mdg4)\(^{02}\) homozygotes are more completely rescued by the third chromosomal transgene (79.0%) as compared to the second cromosomal one (26.6%, Table 1, column 1 and 2). This may be due to chromosomal position effects. On the other hand the rescue ability of the transgene is allele dependent. Higher number of transheterozygous offspring containing the mutant allele mod(mdg4)\(^{neo129}\) was rescued as compared to transheterozygotes containing the mutant allele mod(mdg4)\(^{02}\) (Table 1). The
latter allele is supposed to be a loss of function allele, whereas \( mod(mdg4)^{neo129} \) is an hypomorphic allele (Büchner et al. 2000).

However, the mutant allele \( mod(mdg4)^{07} \) behaves different. Recessive lethality of transheterozygous \( P(w^+ \ Dv \ mod(mdg4) \ 11.5kb \ NotI) \ mod(mdg4)^{neo129}/mod(mdg4)^{07} \) offspring flies is not rescued (0.6% of expected flies hatch) and only a small number of \( mod(mdg4)^{02}/mod(mdg4)^{07} \) offspring is rescued in presence of the second or the third chromosomal \( P(w^+ \ Dv \ mod(mdg4) \ 11.5kb \ NotI) \) transgene, 12.5% and 5.4% respectively. In contrast to most other recessive lethal \( mod(mdg4) \) alleles the mutant allele \( mod(mdg4)^{07} \) is embryonic lethal and was shown to contain substitutions of two conserved amino acids within the BTB domain (D35N and G93S) which are expected to be involved in dimerization (READ et al. 2000). This might indicate an antimorphic nature of this mutant allele.

We did not find significant differences, whether one copy of the transgene is inherited maternally (Table 1) or paternally (data not shown). However, if two copies of the second chromosomal transgene are inherited, one maternal and one paternal, homozygous \( mod(mdg4)^{02} \) flies are completely rescued (106.7% vs. 26.6% in case of one maternal copy).

This indicates dosage dependent effects of the transgene. We also did not observe significant differences in rescue ability of female and male offspring, except in crosses with mutant allele \( mod(mdg4)^{R32} \). In these crosses the number of transheterozygous female offspring is significantly reduced compared to offspring males.

To prove, if mutant rescue solely depends on the expression of the transgene encoded \( D. virilis \) mod(mdg4) isoforms or inter-species mRNA trans-splicing is involved we constructed the \( P(w^+ \ Dv \ mod(mdg4) \ 6.8kb \ NotI-XbaI) \ D. virilis \) transgene. This transgene encodes exclusively the common exons 1 to 4. Therefore functional Mod(mdg4) proteins can only be produced via trans-splicing. Complementation analysis with a second chromosomal transgene revealed that this transgene also partially rescues recessive lethality (Table 1). This
result suggests that proteins encoded from chimeric transcripts facilitate mutant rescue. However, the number of homozygous / transheterozygous offspring flies is significantly reduced compared to the appropriate genotypes containing the \( P(w^+ Dv \text{mod}(mdg4) \ 11.5kb \text{NotI}) \) transgene. Also fertility of female offspring is significantly reduced in presence of the the \( P(w^+ Dv \text{mod}(mdg4) \ 6.8kb \text{NotI-XbaI}) \) D. virilis transgene in contrast to the appropriate females containing the longer transgene.

**Chimeric mod(mdg4) transcripts are generated by mRNA trans-splicing of D. melanogaster and D. virilis pre-mRNAs:**

Recently, we have shown that trans-splicing of mod(mdg4) pre-mRNAs is not restricted to the isoforms encoded by opposite DNA strands but also includes isoforms expected to be produced exclusively by cis-splicing (DORN *et al.* 2001). This suggests that trans-splicing should be the general mechanism to produce all mod(mdg4) transcripts. Additionally the rescue ability of the D. virilis \( P(w^+ Dv \text{mod}(mdg4) \ 6.8kb \text{NotI-XbaI}) \) transgene indicates the generation of functional chimeric mod(mdg4) transcripts.

To experimentally prove the existence of these transcripts we initially performed RT-PCR which allows to differentiate between the chimeric and the corresponding D. virilis transcripts (schematically shown in Figure 4). We chose the most proximal isoform mod(mdg4)-64.2 which is encoded by the same strand as the common exons and deduced a forward primer (ex4-virF), which exclusively hybridizes to common exon 4 of D. virilis mod(mdg4) and a backward primer (64.2-B-RT), hybridizing to the specific exon mod(mdg4)-64.2 of D. melanogaster (cf. Figure 2). The latter primer, despite of three nucleotide substitutions, hybridizes to the orthologous D. virilis exon at the annealing temperature used (cf. MATERIALS AND METHODS). The two orthologous mod(mdg4)-64.2 specific exons contain several nucleotide substitutions within the amplified region.
which includes the position of a single PvuII restriction site (Figure 4A). Restriction of the resulting RT-PCR fragments with PvuII produces an internal 474bp fragment in case of the chimeric D. virilis/D. melanogaster mod(mdg4)-64.2 cDNA (Figure 4B, lane 1) and a 660bp internal fragment in case of D. virilis mod(mdg4)-64.2 cDNA (Figure 4B, lane 2). These two restriction fragments have been used as indicator for the two cDNAs. With an equimolar mixture of both orthologous D. melanogaster mod(mdg4)-64.2 and D. virilis mod(mdg4)-64.2 cDNA clones as template exclusively the D. virilis fragment is detectable after restriction with PvuII (Figure 4B, lane 3). The same result was obtained, if mixed RNA isolated from D. melanogaster and D. virilis was used as RT-PCR template, indicating the absence of significant template switching.

Chimeric transcripts could be detected if RNA from females containing one copy of the second chromosomal or the third chromosomal P(w+ Dv mod(mdg4) 11.5kb NotI) transgene is used as template for RT-PCR (Figure 4B, lanes 5 and 6, respectively). Also in 3-P(w+ Dv mod(mdg4) 11.5kb NotI) mod(mdg4)neo129 homozygotes and in 2-P(w+ Dv mod(mdg4) 11.5kb NotI)/+; mod(mdg4)02/mod(mdg4)02 the chimeric transcript mod(mdg4)-64.2 is clearly detectable (Figure 4B, lane 7 and 8, respectively). The increased ratio of the chimeric 474bp fragment in females containing two copies of the transgene (Figure 4B, lane 7) indicates dosage dependent accumulation of the chimeric transcript. The smaller chimeric-specific fragment of 311bp is not visible in most lanes because of the underrepresentation of the corresponding amplicon. However, if an equimolar mixture of D. virilis and chimeric D. vir./D. mel. mod(mdg4)-64.2 cDNA clones is used as template all expected fragments, including the 311bp fragment, are clearly visible (Figure 4B, lane 9).

Next we proved if isoforms not encoded by the D. virilis transgene are produced as chimeric transcripts. We used the D. virilis specific forward primer ex4-vir-F and three D. melanogaster specific backward primers, mod(mdg4)-51.4-RT-B, mod(mdg4)-56.3-RT-B
and mod(mdg4)-67.2-RT-B. In three independent RT-PCR experiments with RNA from females of the genotype 2-P(w+ Dv mod(mdg4) 11.5kb NotI)/+; mod(mdg4)02/mod(mdg4)02 fragments of the expected size have been obtained and sequencing revealed the expected chimeric cDNAs. This result suggests, that obviously all Mod(mdg4) isoforms can be produced by inter-species trans-splicing. We conclude, that important features involved in trans-splicing of mod(mdg4) are evolutionary conserved between D. melanogaster and D. virilis.

The semiquantitative RT-PCR experiments described above do not allow an accurate estimation of the frequency of inter-species mRNA trans-splicing. Therefore we cloned the resulting mod(mdg4)-64.2 RT-PCR fragments obtained from homozygous 3-P(w+ Dv mod(mdg4) 11.5kb NotI) females in three independent experiments in pGEM-T and tested altogether 282 individual clones for their identity via specific primer pairs (cf. MATERIALS AND METHODS). Three of these clones were proved to be chimeric whereas the remaining 279 represent D. virilis cDNA clones. All chimeric clones and 59 of the D. virilis clones have been confirmed by sequencing. Also in these experiments no sign of template switching was found. According to these results the chimeric transcripts comprise about one percent of the D. virilis mod(mdg4)-64.2 transcript encoded by the transgene. To confirm these results we next performed real-time RT-PCR experiments.

First we determined the expression level of D. melanogaster mod(mdg4) with respect to ribosomal protein 49 mRNA (rp49) with a specific primer pair deduced from mod(mdg4) common exons using total RNA isolated from w1118 females as control. According to these results mod(mdg4) expression is 76 fold higher than rp49 (Table 2). The two specific D. melanogaster isoforms mod(mdg4)-64.2 and mod(mdg4)-67.2 are represented with less than 1% each with respect to mod(mdg4) common exons. It has to be noted, that expression of these isoforms is still in the range of rp49 expression. At least mod(mdg4)-67.2 is supposed
to be one of the most abundant mod(mdg4) isoforms (BÜCHNER et al. 2000; GERASIMOVA et al. 1995) and therefore expected at significant higher expression levels compared to most other mod(mdg4) isoforms. However, if trans-splicing is the main mechanism for generating all mod(mdg4) isoforms (DORN and KRAUSS 2003), the transcript containing common exons 1 to 4 which is used as splice donor for all isoforms should be expressed at high level.

Next we used RNA isolated from w^{1118} females containing two copies of the third chromosomal P(w^+ Dv mod(mdg4) 11.5kb NotI) transgene as template for real-time qRT-PCR. The endogenous D. melanogaster common exons are expressed at almost the same level (75.6, Table 2) as in the control females, whereas the transgenic D. virilis common exons (59.2) are expressed at slightly lower level (Table 2). In these transgenic females the isoform mod(mdg4)-64.2 is expressed from the endogenous mod(mdg4) locus and from the D. virilis transgene. Whereas expression of the D. melanogaster isoform is decreased by factor 2 compared to control females (0.314 versus 0.667, respectively) the transgenic D. virilis orthologous isoform is increased by factor 2 (1.2). Isoform mod(mdg4)-67.2 is not encoded by the D. virilis transgene. Also in this case expression is decreased in females with two copies of the P(w^+ Dv mod(mdg4) 11.5kb NotI) transgene (0.451 versus 0.654). It is not clear whether these differences reflect changes in general transcriptional activity in both genotypes or feed back mechanisms regulating the mod(mdg4) expression are involved.

To determine the ratio of chimeric transcripts specific primer pairs have been deduced. Real-time RT-PCR experiments indicate that chimeric D. virilis/D. melanogaster mod(mdg4)-64.2 represents 2.5% with respect to the corresponding endogenous D. melanogaster isoform (0.008/0.314, Table 2), whereas the chimeric mod(mdg4)-67.2 isoform represents 4.7% of the corresponding endogenous transcript (0.021/0.451, Table 2). These results indicate, that
chimeric transcripts are produced at significant level. The frequency of chimeric mod(mdg4)-64.2 is in similar range as determined by analysing individual RT-PCR clones.

Next we performed the same RT-PCR experiments using RNA isolated from $w^{118}, P(w^+ Dv\ mod(mdg4) 6.8kb NotI-XbaI) / P(w^+ Dv\ mod(mdg4) 6.8kb NotI-XbaI)$ females. Expression of D. virilis and the endogenous common exons as well as endogenous isoforms mod(mdg4)-64.2 and –67.2 is decreased compared to $P(w^+ Dv\ mod(mdg4) 11.5kb NotI)$ transgenic females. The ratio of the two corresponding chimeric transcripts is increased approximately two fold (3.8% and 11.7%, respectively). This increase could be explained by the absence of the D. virilis specific exons in the short transgene trapping a significant fraction of the common exons as splice donor.

We conclude from these experiments, that proteins produced from chimeric mod(mdg4) transcripts in transgenic $P(w^+ Dv\ mod(mdg4) 6.8kb NotI-XbaI)$ flies are sufficient to rescue viability at least partially. The improved rescue ability of $P(w^+ Dv\ mod(mdg4) 11.5kb NotI)$ transgene can thus be explained by the additional expression of the five proximal orthologous D. virilis Mod(mdg4) isoforms encoded by this transgene.

Mod(mdg4) staining pattern of polytene chromosomes of mutant larvae is restored in presence of the D. virilis transgene:

Previously, we demonstrated that Mod(mdg4) proteins are not detected on salivary gland polytene chromosomes of homozygous mod(mdg4)$^{02}$ third larval stages (BUCHNER et al. 2000). To prove, if in the presence of the D. virilis transgene binding of Mod(mdg4) proteins is restored we performed immunostaining of 2-$P(w^+ Dv\ mod(mdg4) 11.5kb NotI)/+$; mod(mdg4)$^{02}$/mod(mdg4)$^{02}$ larvae with anti-Mod(mdg4)-58.0$^{\text{BTB-534}}$, an antiserum detecting all Mod(mdg4) isoforms. Independent experiments clearly indicate significant immunostaining in the presence of the transgene (Figure 5, B). Similar results were obtained
for other transheterozygotes. Next we used the specific antibody anti-Mod(mdg4)-67.2\textsuperscript{403-610}, detecting exclusively this isoform. The staining pattern of polytene chromosomes of 2-\textit{P}(\textit{w}^+ \textit{Dv mod(mdg4) 11.5kb NotI}; mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} larvae is similar to that of \textit{mod(mdg4)}\textsuperscript{+} chromosomes (Figure 5, E and H), indicating that significant levels of chimeric Mod(mdg4)-67.2 protein is produced. In \textit{mod(mdg4)}\textsuperscript{02} homozygous larvae no staining is detected (Figure 5L). To prove if all Mod(mdg4)-67.2 binding sites detected in wildtype larvae are also found in transgenic 2-\textit{P}(\textit{w}^+ \textit{Dv mod(mdg4) 11.5kb NotI}); 

\textit{mod(mdg4)}\textsuperscript{02}/\textit{mod(mdg4)}\textsuperscript{02} larvae we analyzed a selected region on the X chromosome by comparing the appropriate staining pattern (Figure 5, N and O). At least in this region all strong binding sites correspond to each other indicating both, the re-establishment of the staining pattern of Mod(mdg4)-67.2 and the binding specificity of the chimeric protein in transgenic larvae. However, using the antibody detecting all isoforms the staining pattern and the signal intensity was more variable. We were not able to perform staining of polytene chromosomes of 2-\textit{P}(\textit{w}^+ \textit{Dv mod(mdg4) 6.8kb NotI-XbaI}/+; mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} larvae, because salivary glands and nuclei were reduced in size and changed chromosome morphology prevented reproducible antibody staining. To prove if the specific antibody detects the expected full length chimeric Mod(mdg4)-67.2 protein we performed Western analysis with salivary glands of third instar 2-\textit{P}(\textit{w}^+ \textit{Dv mod(mdg4) 11.5kb NotI}); 

\textit{mod(mdg4)}\textsuperscript{02}/\textit{mod(mdg4)}\textsuperscript{02} and 2-\textit{P}(\textit{w}^+ \textit{Dv mod(mdg4) 6.8kb NotI-XbaI}/+; 

\textit{mod(mdg4)}\textsuperscript{02}/\textit{mod(mdg4)}\textsuperscript{02} larvae (Figure 6). A protein with the same molecular weight as in isogenic \textit{w}\textsuperscript{1118} larvae as control for \textit{mod(mdg4)}\textsuperscript{+} (Figure 6, lane 1) could be clearly detected in both transgenic \textit{mod(mdg4)}\textsuperscript{02}/\textit{mod(mdg4)}\textsuperscript{02} genotypes (Figure 6, lanes 3 and 4) whereas in \textit{mod(mdg4)}\textsuperscript{02} homozygotes the protein is absent (Figure 6, lane 2). The protein level is decreased in transgenic mutant larvae. However, our results indicate, that proteins encoded
by chimeric transcripts which are generated from pre-mRNAs transcribed from nonhomologous chromosomes can be produced in significant amount.
DISCUSSION

The limited knowledge on the functional significance of the large number of mod(mdg4) isoforms and the unusual type of gene structure in *D. melanogaster* prompted us to analyze the orthologous locus from the distantly related species *D. virilis*. It represents an evolutionary distant species that has been separated about 40-60 million years ago from the *Sophophora* which includes *D. melanogaster*. This period of time allows the selection for functionally essential genes. A number of orthologous genes have been studied in detail and their functional conservation in *D. virilis* was demonstrated by mutant rescue experiments (KASSIS *et al.* 1986; COLOT *et al.* 1988; HART *et al.* 1993; BOPP *et al.* 1996). The degree of the overall conservation within coding regions is variable and can reach up to 98% similarity (TOMINAGA *et al.* 1992). Our results demonstrate a strong evolutionary conservation of all Mod(mdg4) isoforms identified in *D. virilis*, indicating the functional significance of the multiple isoforms. We have previously presented evidence for a functional differentiation of at least two isoforms, Mod(mdg4)-58.0 and Mod(mdg4)-67.2 in *D. melanogaster* (BÜCHNER *et al.* 2000). The high degree of sequence conservation of both isoforms in *D. virilis* is in good agreement with binding to corresponding sites on polytene chromosomes as shown for isoform Mod(mdg4)-58.0. Its binding to corresponding subdivisions on polytene chromosomes suggests an involvement in regulation of a subset of orthologous genes in *D. melanogaster* and *D. virilis*.

The common N-terminal region which is the part of all isoforms and therefore supposed to contribute general functions shows an extended identity beyond the BTB/POZ domain. This common protein region represents about two third of any of the Mod(mdg4) proteins. GAUSE *et al.* (2001) have shown, that the ubiquitously expressed protein Chip interacts with the common region of Mod(mdg4) in *D. melanogaster*. Chip is supposed to facilitate
enhancer-promoter interactions in a large number of genes and was shown to interact genetically and physically with several LIM- and homeodomain containing transcription factors (MORCILLO et al. 1997; TORIGOI et al. 2000; HEITZLER et al. 2003). These data together with the observed pleiotropic mutant effects of most mod(mdg4) mutants indicate a putative link between the several hundred binding sites of Mod(mdg4) on polytene chromosomes and their involvement in transcriptional regulation of a large number of genes. The strong conservation of the common protein region in both Drosophila species might be the consequence of the evolutionary conserved interaction with Chip and other putative interacting proteins. The N-terminal BTB/POZ domain is almost identical in both species. This domain was shown to mediate homo- and/or heterodimerization (BARDWELL and TREISMAN 1994). A similar degree of conservation between D. melanogaster and D. virilis was found for the BTB/POZ domain containing gene GAGA/Trl (LINTERMANN et al. 1998). Also in this case at least two alternatively spliced isoforms containing a common N-terminal region of 400 amino acids but variable C-termini have been described. However, in contrast to mod(mdg4) no significant functional differentiation between the two GAGA isoforms has been described (SOELLER et al. 1993; BENYAJATI et al. 1997). If specific C-termini of orthologous Mod(mdg4) isoforms are compared, a remarkable degree of identity within the FLYWCH domain, a Cys$_2$His$_2$ motif containing protein domain, is found. This domain is supposed to be involved in protein-protein interactions (DORN and KRAUSS 2003 and references therein). Strong conservation of most amino acid positions within this motif between orthologous isoforms implies their functional importance for isoform specific interactions with other proteins. The unique C-terminal region of isoform Mod(mdg4)-67.2 has been demonstrated to interact with Su(Hw) to create a functional gypsy insulator element (GAUSE et al. 2001; GHOSH et al. 2001) whereas the unique C-terminus of isoform Mod(mdg4)-56.3/Doom interacts with Baculovirus Inhibitor of apoptosis
protein/IAP (HARVEY et al. 1997). The high degree of sequence identity suggests, that these interactions are conserved in *D. virilis*. If the orthologous *D. virilis* isoforms Mod(mdg4)-64.2, Mod(mdg4)-60.1 and Mod(mdg4)-67.2 are compared with their counterparts in *D. melanogaster*, it becomes evident that additional amino acid positions flanking the FLYWCH motif are highly conserved. However, the extension and the location of the identity beyond the FLYWCH motif is isoform dependent. In case of Mod(mdg4)-67.2 an additional strongly conserved sequence motif of 22 amino acids is located at the C-terminus. Based on pull down experiments with a C-terminal truncated (deletion of 43 amino acids) Mod(mdg4)-67.2 protein and the observed phenotype connected with the corresponding mutant protein (Mod(mdg4)-67.2<sup>T6</sup>) the FLYWCH domain itself is not sufficient for interaction with Su(Hw) (GAUSE et al. 2001), indicating the functional importance of the strongly conserved 22 C-terminal amino acids. Also, the isoforms without the FLYWCH motif are conserved as shown for Mod(mdg4)-58.0 (identity of 51% within the unique C-terminus). Recently, an evolutionary analysis of several Dipteran orthologous *mod(mdg4)* loci revealed a significant conservation of most isoforms including Mod(mdg4)-58.0, Mod(mdg4)-60.1, Mod(mdg4)-64.2 and Mod(mdg4)-67.2 (LABRADOR and CORCES 2003; KRAUSS and DORN 2004).

Two conclusions can be drawn from the evolutionary conservation of Mod(mdg4) proteins. First, the large number of isoforms is functionally important in both Drosophila species and second, the conservation of the unique C-terminal regions clearly points to a functional differentiation between single isoforms.

In the present study we demonstrate for the first time that along with the evolutionary conservation of the unusual gene structure of *mod(mdg4)* in *D. virilis* mRNA trans-splicing is also conserved in both species. We performed three different assays to prove the existence of chimeric transcripts *in vivo*. The identification of chimeric *mod(mdg4)* isoforms in
transgenic flies clearly indicates, that the mechanism of mRNA trans-splicing is conserved between the distantly related Drosophila species. Quantitative RT-PCR experiments reveal that in case of isoform Mod(mdg4)-67.2 the chimeric *D. virilis/D. melanogaster* transcript in transgenic flies containing two copies of the second chromosomal *P(w+ Dv mod(mdg4) 6.8kb NotI-XbaI*) transgene represents about 12% of the corresponding endogenous *D. melanogaster* transcript. The Mod(mdg4)-67.2 protein can be clearly detected on polytene chromosomes of 2-*P(w+ Dv mod(mdg4) 11.5kb NotI)/+; mod(mdg4)02/mod(mdg4)02* larvae but not in *mod(mdg4)02* homozygous larvae. Because the specific *mod(mdg4)*-67.2 exons are not encoded by the *D. virilis* transgene, this result strongly suggests, that the cytologically detected protein represents the chimeric *D. virilis/D. melanogaster* Mod(mdg4)-67.2 protein, which is produced in significant amount. In fact the presence of considerable amounts of the full length Mod(mdg4)-67.2 protein was demonstrated in Western blot analysis. The maintenance of the binding pattern of the chimeric Mod(mdg4)-67.2 isoform compared to the *D. melanogaster* Mod(mdg4)-67.2 on polytene chromosomes also implicates the functional conservation of the *D. virilis* N-terminal region.

Recently, MONGELARD et al. (2002) demonstrated that interallelic complementation is facilitated by mRNA trans-splicing if two mutations disrupting independent *mod(mdg4)* mRNAs are combined in trans. They assume that the close proximity of donor and acceptor mRNAs within the *mod(mdg4)* locus is a prerequisite for generation of significant amounts of wildtype Mod(mdg4)-67.2 protein. The *lola* locus of *D. melanogaster* represents a second complex gene where mRNA trans-splicing was demonstrated (HORIUCHI et al. 2003). Mutations interfering with the pairing of the *lola* locus reduce the in vivo trans-splicing of isoform T from 44% to 1%. However, the authors did not prove the consequences on protein level. Our transgene assay clearly demonstrates that even underrepresented chimeric transcripts produced from mRNAs encoded by nonhomologous chromosomes can produce
considerable levels of the corresponding protein. Mutant rescue experiments with two
different D. virilis mod(mdg4) transgenes indicate the functional conservation of Mod(mdg4)
protein isoforms. Both, the P(w+ Dv mod(mdg4) 11.5kb NotI) transgene which encodes the 5
proximal isoforms and the P(w+ Dv mod(mdg4) 6.8kb NotI-XbaI) transgene encoding
exclusively common exons 1 to 4 facilitate rescue of recessive lethality of mod(mdg4) mutant
alleles. We suppose, that the rescue ability of the short transgene depends mainly on its
capacity to produce sufficient chimeric transcripts consisting of the D. virilis common exons
and the endogenous D. melanogaster specific exons, which was demonstrated at least for
isoform Mod(mdg4)-67.2. However, the significantly reduced rescue ability of the shorter
transgene indicates, that all / some isoforms have to exceed a critical threshold to restore
viability completely. The P(w+ Dv mod(mdg4) 11.5kb NotI) transgene, which produces 5
orthologous D. virilis isoforms significantly improves rescue ability. We can not exclude
position effects influencing the expression level of the transgene. Further experiments with a
series of independent insertions of the short transgene scattered throughgout the genome
should provide further insight into a putative correlation of genomic transgene position and
efficiency of trans-splicing.
The observed frequency of chimeric transcripts, although significantly lower as compared to
the corresponding endogenous transcript can be interpreted in two ways. First, the splice
donor containing the D. virilis mod(mdg4) common exons is produced at high level enabling
its spreading in the nucleus. Thus a significant number of donor molecules gets in close
proximity to mod(mdg4) acceptor mRNAs, even if they are transcribed from a
nonhomologous chromosome. The much higher expression of the common exons compared
to the specific isoform mod(mdg4)-67.2 in w^{1118} females (116 fold, cf. Table 2) is in
agreement with this hypothesis. A second explanation supposes transcription of both
precursor mRNAs within the same compartment of the nucleus (for review see, COCKEL and GASSER 1999) and thereby increasing the frequency of chimeric mRNAs.

**Acknowledgement:** We would like to thank P. Geyer, T. Parnell and B. Gilmore for the generous gift of the specific antibody anti-Mod(mdg4)-67.2403-610, Anja Ebert for help in immunocytology, M. Kube for technical assistance and K. Kittlaus for sequence analysis. We are indebted to S. Phalke for critical reading of the manuscript. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to R.D. (Do407/2-2).
LITERATURE CITED


GHOSH, D., T. I. GERASIMOVA and V. G. CORCES, 2001 Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. EMBO J. **20**: 2518-2527.


SPRADLING, A. C. et al., 1998  The BDGP gene disruption project: Single P-element insertions mutating 30% of *Drosophila* autosomal genes, Flybase.


### TABLE 1

Complementation analysis of mod(mdg4) alleles in presence of *D. vir mod(mdg4)* transgenes. The \( P(w^+ \textbf{Dv mod(mdg4)} \ 11.5\text{kb } \text{NotI}) \) transgene encodes common exons 1 to 4 and the most proximal specific exons \( \text{mod(mdg4)}-64.2, \text{mod(mdg4)}-60.1, \text{mod(mdg4)}-53.5, \text{mod(mdg4)}-55.1 \) and \( \text{mod(mdg4)}-53.1 \), whereas the transgene \( P(w^+ \textbf{Dv mod(mdg4)} \ 6.8\text{kb } \text{NotI-XbaI}) \) encodes exclusively the common exons 1 to 4.

<table>
<thead>
<tr>
<th>genotype/ males</th>
<th>females</th>
<th>chromosomal location of the transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
<td>sl (26.6%/673)</td>
<td>( P(w^+ \textbf{Dv mod(mdg4)} \ 11.5\text{kb})/++; \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
</tr>
<tr>
<td>( \text{mod(mdg4)}^{neo129}/\text{TM3,Sb Ser} )</td>
<td>sv (95.1%/687)</td>
<td>( P(w^+ \textbf{Dv mod(mdg4)} \ 11.5\text{kb})/++; \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
</tr>
<tr>
<td>( \text{mod(mdg4)}^{R32}/\text{TM3,Sb Ser} )</td>
<td>sl/( \text{sv}^m ) (26.2%/91.8%/765)</td>
<td>( P(w^+ \textbf{Dv mod(mdg4)} \ 11.5\text{kb})/++; \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
</tr>
<tr>
<td>( \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
<td>sl (12.5%/635)</td>
<td>( P(w^+ \textbf{Dv mod(mdg4)} \ 6.8\text{kb})/++; \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
</tr>
<tr>
<td>( P(w^+ \textbf{Dv mod(mdg4)} \ 11.5\text{kb})/++; \text{mod(mdg4)}^{62}/\text{TM3,Sb Ser} )</td>
<td>+ (106.7%/1097)</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

For viability: + complementation; sv – subvital, on average less than 100% and more than 50% of the expected progeny survive to eclose as adults; sl – semilethal, on average 5-50% of the expected progeny survive to eclose as adults; - no complementation, less than 5% of the expected progeny survive to eclose as adults; the percentage of expected homo-, trans-heterozygous offspring / total number of flies is shown in brackets; * viability of transheterozygous offspring females (f) and males (m) is shown.
TABLE 2

Results of the real-time quantitative RT-PCR. For this experiment RNA was extracted from adult females of $w^{1118}$, $w^{1118}$; 3-$P(w^+ Dv\ mod(mdg4)\ 11.5kb\ NotI)/3-P(w^+ Dv\ mod(mdg4)\ 11.5kb\ NotI)$ and $w^{1118}$; 2-$P(w^+ Dv\ mod(mdg4)\ 6.8kb\ NotI-XbaI)/2-P(w^+ Dv\ mod(mdg4)\ 6.8kb\ NotI-XbaI)$.

<table>
<thead>
<tr>
<th>genotype</th>
<th>relative expression (normalized to $rp49^*$)</th>
<th>$w^{1118}$ (control)</th>
<th>3-$P(w^+ Dv\ 11.5kb)$</th>
<th>2-$P(w^+ Dv\ 6.8kb)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>common exons 1-4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D.\ mel.$</td>
<td>76.0 ± 12.1</td>
<td>75.6 ± 13.8</td>
<td>44.2 ± 5.4</td>
<td></td>
</tr>
<tr>
<td>$D.\ vir.$</td>
<td>–</td>
<td>59.2 ± 12.7</td>
<td>27.8 ± 6.2</td>
<td></td>
</tr>
<tr>
<td><strong>specific isoforms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D.\ mel.\ 64.2$</td>
<td>0.667 ± 0.042</td>
<td>0.314 ± 0.044</td>
<td>0.132 ± 0.010</td>
<td></td>
</tr>
<tr>
<td>$D.\ vir.\ 64.2$ (transgenic)</td>
<td>–</td>
<td>1.200 ± 0.178</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>$D.\ vir./D.\ mel.\ 64.2$ (chimeric)</td>
<td>–</td>
<td>0.008 ± 0.001</td>
<td>0.005 ± &gt;0.001</td>
<td></td>
</tr>
<tr>
<td>$D.\ mel.\ 67.2$</td>
<td>0.654 ± 0.085</td>
<td>0.451 ± 0.051</td>
<td>0.214 ± 0.013</td>
<td></td>
</tr>
<tr>
<td>$D.\ vir./D.\ mel.\ 67.2$ (chimeric)</td>
<td>–</td>
<td>0.021 ± 0.004</td>
<td>0.025 ± 0.003</td>
<td></td>
</tr>
</tbody>
</table>

*$rp49=1$
FIGURE 1
Distribution of Mod(mdg4) protein isoforms on salivary gland polytene chromosomes of *Drosophila virilis*. (A) The polyclonal antiserum anti-Mod(mdg4)-58.0$^{403-534}$ specifically recognizing Mod(mdg4)-58.0 protein isoform in *D. melanogaster* shows a similar staining pattern in *D. virilis*. (B) Localization of Mod(mdg4) at the BX-Complex in *D. melanogaster* and *D. virilis* by sequential *in situ* hybridization with a *Ubx*-DNA probe and antibody staining with anti-Mod(mdg4)-58.0$^{BTB-534}$. In both species the binding of Mod(mdg4) is close to *Ubx*.

FIGURE 2
Comparison of genomic structure of *mod(mdg4)* of *D. melanogaster* and *D. virilis*. Above the scale the *mod(mdg4)* exon/intron structure of *D. melanogaster* and *D. virilis* is shown. Translated regions are represented as rectangles (black, encoded by the same strand as common exons, grey, encoded by the complementary strand). Non translated regions are shown as bars. The alternatively spliced specific exons and their corresponding orthologs in *D. virilis* are indicated by numbers representing the deduced molecular weight of the *D. melanogaster* isoforms. The alternatively used splice site of exon 4 is indicated by multiple bars. Primers used in RT-PCR experiments are indicated by arrows. Genomic fragments tested for mutant rescue are indicated below the scale.

FIGURE 3
Sequence comparison of orthologous Mod(mdg4) protein isoforms of *D. melanogaster* and *D. virilis*. (A) Alignment of the common N-terminal protein sequences encoded by exons 2 to 4 from *D. melanogaster* and *D. virilis mod(mdg4)*. The BTB domain is boxed (dark grey). (B) Comparison of the putative specific C-terminal protein sequences of the isoforms.
Mod(mdg4)-64.2, Mod(mdg4)-60.1, Mod(mdg4)-67.2, Mod(mdg4)-55.1 and Mod(mdg4)-58.0. Identical amino acid positions are indicated by stars, double dots and single dots indicate functional or structural similar amino acids. The FLYWCH motif is indicated by a box (light grey) and the strongly conserved amino acid positions within the FLYWCH motif are in bold.

FIGURE 4
RT-PCR analysis demonstrating the expression of *D. virilis mod(mdg4)-64.2* and the chimeric *D. virilis/D. melanogaster mod(mdg4)-64.2* transcript in transgenic *D. melanogaster* females. (A) Schematic representation of the RT-PCR assay to determine the ratio of both transcripts. In the upper part relevant *mod(mdg4)* transcripts (wavy bars, TSS represents the *trans*-splice site) encoded from the endogenous *D. melanogaster* (Dm) locus (red bar) and from the *D. virilis* (Dv) transgene (black bar) are shown. Snapped off arrows indicate independent initiation of transcription. Below the bold arrow cDNA fragments obtained after RT-PCR (filled bars) and restriction fragments resulting after digestion with the endonuclease *Pvu*II (thin lines) are shown. For RT-PCR a forward primer specific for *D. virilis* exon 4 and a backward primer annealing to exon 5 of *mod(mdg4)-64.2* of both species were used. (B) Results of the RT-PCR to demonstrate the expression of the *D. virilis* and the chimeric *D. virilis/D. melanogaster mod(mdg4)-64.2* isoforms. Templates used for PCR are: lane 1 – chimeric cDNA clone *D. vir./D. mel. mod(mdg4)-64.2*, lane 2 – cDNA clone *D. vir. mod(mdg4)-64.2*, lane 3 – equimolar mixture of *D. mel. mod(mdg4)-64.2* and *D. vir. mod(mdg4)-64.2* cDNA clones, lane 4 – 100bp ladder, lanes 5 and 6 – RT templates of transgenic females with the genotype 2-P(*w*+/Dv *mod(mdg4) 11.5kb NotI*)/+ and 3-P(*w*+/Dv *mod(mdg4) 11.5kb NotI*)/+, lanes 7 and 8 – 2-P(*w*+/Dv *mod(mdg4) 11.5kb NotI*)/*mod(mdg4)neo129* homozygotes and 3-P(*w*+/Dv *mod(mdg4) 11.5kb NotI*)/+;
mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02}, lane 9 – mixture of D. vir. mod(mdg4)-64.2 cDNA and chimeric D. vir./D. mel. mod(mdg4)-64.2 cDNA in a ratio of 1:1.

**FIGURE 5**

The *D. virilis* mod(mdg4) transgene partially rescues the Mod(mdg4) staining pattern on polytene chromosomes of homozygous mod(mdg4) mutant larvae.

Staining of polytene chromosomes of third instar larvae with genotype 2-\textit{P}(w\textsuperscript{+} D\textit{v} mod(mdg4) 11.5kb NotI); mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} (A,B,C), \textit{w}\textsuperscript{1118} (D,E,F), 2-\textit{P}(w\textsuperscript{+} D\textit{v} mod(mdg4) 11.5kb NotI); mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} (G,H,J) and mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} larvae not containing the transgene (K,L,M). A,D,G,K – propidium iodid staining, B – staining with anti-Mod(mdg4)-58.0\textsuperscript{BTB-534}, E,H,L – staining with anti-Mod(mdg4)-67.2\textsuperscript{403-610}, C,F,J,M – merged images.

(N) Comparison of the staining pattern of the tip of the X chromosome of \textit{w}\textsuperscript{1118} and 2-\textit{P}(w\textsuperscript{+} D\textit{v} mod(mdg4) 11.5kb NotI); mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02} larvae.

**FIGURE 6**

Western blot analysis of dissected salivary glands with anti-Mod(mdg4)-67.2\textsuperscript{403-610} antibody. lane 1 – \textit{w}\textsuperscript{1118}, lane 2 – mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02}, lane 3 – 2-\textit{P}(w\textsuperscript{+} D\textit{v} mod(mdg4) 11.5kb NotI); mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02}, lane 4 – 3-\textit{P}(w\textsuperscript{+} D\textit{v} mod(mdg4) 6.8kb NotI-XbaI)/+; mod(mdg4)\textsuperscript{02}/mod(mdg4)\textsuperscript{02}. anti-Tubulin antibody staining was used as loading control.
(A) Common region of Mod(mdg4) protein isoforms

D. mel  
MADDEQFSLCWNNFNNTLNSAGFHESLRCGLDLDVSLAAEGQIVKAHRVLTVSCHSPFRFKM  
D. vir  
MADDEQFSCLWNFSNNTLNSAGFHESLRCGLDLDVSLAAEGQIVKAHRVLTVSCHSPFRFKM

D. mel  
FTQMPSNTHAVIFLNVSHSLKLDLIFQMYCGEHNVIDQDHALPAFISTASLQIGITLDND  
D. vir  
FTQMPSNTHAVIFLNVSHSLKLDLIFQMYCGEHNVIDQDHALPAFISTASLQIGITLDND

D. mel  
PAPQPQESPPPPAAPHVQQQ---IPAAVRQRQQRASARNKLYETVDDGLGDEKQSTTQ  
D. vir  
PAPQPQEPPTPPPPAAPHVQQQQQVDPAAVRQRQQRASARNKLYETVDDGLGDEKQSTTQ

D. mel  
IVIQTTAAPQATIVQQQQ--PQQAAQQIQSQQLQTGTTTTATLVSNKRSACQRTPTTAS  
D. vir  
IVIQTTAAPQATIVQQQQQQHTGQQTATTTATLVSNKRSACQRTPTTAS

D. mel  
SSAGVKSRTSUANMVDPLDTSTT---GATTTAQLVPQQITVQTSVBBAAAAAAAAA  
D. vir  
---GAVKRSTSUANMVDPLDTSTT---GATTTAQLVPQQITVQTSVBBAAAAAAAAA

D. mel  
SYFTENEDATGNAQTHAANTSGG---GVTATTSKAVVKQQSQSYSDSFVDTGAEQSNTEA  
D. vir  
SYFTENEDATGNAQTHAANTSGG---GVTATTSKAVVKQQSQSYSDSFVDTGAEQSNTEA

(B) Specific regions of Mod(mdg4) protein isoforms

Mod(mdg4)-64.2

D. mel  
QDTEIS
F
FRISQKKNAQ
LVFRNYI
Y
NKKLTQANGQTT
W
R
C
ADVLKLR
C
KAVVITRDGHFID  
D. vir  
--GKIQ
F
FRISQKKNAQ
LVYRDYI
Y
NKKLTQANGQTT
W
R
C
ADVLKLR
C
KAVVITKNGEFLD

D. mel  
ARRQ
H
H
ESHASRIGQRQLYKVEQELEEYIEICTSNPKISQYLGSSNI-IVTAKDGKDCK  
D. vir  
ARRQ
H
H
ESHASRIGQRQLYNVEEELDEYIEICTSNPKISQYLSSSNISIMTAKDGKECK

D. mel  
LFLPAAEATEIEMQALVDAAEEELDEEERHAEERIRDRQGWRTEEAKHRSLKSEHP  
D. vir  
LYVPASEATEIEMQALVDAADDEED-----LMERVVELKPIQRLSHKLKL

Mod(mdg4)-60.1

D. mel  
EDELVFIESPSTPCVLNMGYMNCHSRKSNKQYWRCHYNKKAHEMCRSVCRLNRL  
D. vir  
NQCLLFIESPSTPCVLNMGYMNCHSRKSNKEYWRCHYNKKAHEMCRSVCRLNRL

D. mel  
KSITGGLHNPHTHEDGKIQRNKMAIGG-TKLSRTSHFTQLQLO-- --  
D. vir  
KSITGGLHNPHTHEDGKIQRNLRAGVSTAHKLKLRHSYQPQKEEAEABEQ

D. mel  
---EQQEQ--FIDEHOLTSDAATLQLTDQELIHLSMMHLHE  
D. vir  
QQQMLLTEQQEQPQSMLATDAAATLQLTEHLSMMLH

Mod(mdg4)-55.1

D. mel  
-----------GLIFK--AARHIAPIQKVRQ--V--R-----DKFL  
D. vir  
YTYSQRTNRNMSPLLRDRDFQPVIFPRTSIKQLSATAVKRRTPATNLNVDAYV
D. mel   ATIIKLEPAGRLNKNPDIIRTSSNEHNFVYVGLPRMKGCNCLKKNRTGLRRINTLC  85
D. vir   AVSQMEPVVRALNLSNPENILRTSSQHNFVVFVGVPRMKGCNTCLKGLGKTNLRLTTCC

D. mel   NTCPGSNMMCEPCFEELHS  103
D. vir   DTCPSNMMCEPCFEELH-

Mod(mdg4)-58.0

D. mel   DGPSKDTAIPKPAEHPRKPATDSVQKSRSVDAIPLPDGSRVFSKVALAKAYIPPMIY  60
D. vir   -VPPK----QTPTGKGQKEAEISI-KATKDAR-IPLFGGSKVFVSKEDLVYVPIPPFY

D. mel   TCRVMDLVIKDLVRIAGHEETTDKDLIQDIITHVCVFAQLRGQLTPSAVQEFIDHKL 120
D. vir   TSRVTDLLIEKLGELQPEKDLFQAIISHVCVQVFARRGDKLEATVQEFIDHKL

D. mel   STLKLMPKIEGK  132
D. vir   ATIKLFPCMK--  124

Mod(mdg4)-67.2

D. mel   --QAATSASATKIPPRKRGRPPTKVEDQTPPKLLEKLQAATLNEASEPAVYASTTKG  58
D. vir   APATZTLPAQVIPPRKRPRTPKVEDQPAKLLEKLQAATLNEASEPAVYASTTKG

D. mel   VRLIFNHHLFKSFFKADYSVFQCCYREHGEECKVRFVCDQKRFPYGEHVFQASDK 118
D. vir   VRLLVYEHFLFKSFKAEFSVFQCCFREHGEECKVRFVCDQKNVFPDGEHVFQASDK

D. mel   SCLPSQFMPGESGVISSLPS--KELLMNNTKLEADDKEDDFEEFEIQEIDEIHEL 175
D. vir   SVTSSQFMPGESAVSPKEFKEKQEVIQKSVAKVDGDEEEEEEVEFEEIHEIHELADAN

D. mel   EPE---------KTPAK-EEE----------VDPNDFREIKRRLQKALQNKKK 210
D. vir   ESTGTETADGVLAPAGEEEMSNIKPTGEIDPNDPFEKIRRLQKALQNKKK 235
A

endogenous Dm mod(mdg4) specific exons 5

mRNA

5' TSS 5'

Dv mod(mdg4) transgene common exons 1-4 specific exons 5

RT-PCR restriction with Pvu II

Dv cDNA

78bp 660bp 128bp

Dv/Dm chimeric cDNA

78bp 474bp 311bp

B

1 2 3 4 5 6 7 8 9

660bp 474bp 311bp 128bp 78bp
Mod(mdg4)-67.2^{403-610} antibody

α-Tubulin