



ELSEVIER

Gene 215 (1998) 93–100

GENEAN INTERNATIONAL JOURNAL ON
GENES AND GENOMES

Evolutionary conservation of a testes-specific proteasome subunit gene in *Drosophila*

John M. Belote *, Mary Miller, Kerrie Ann Smyth

Department of Biology, Syracuse University, 130 College Place, Syracuse, NY 13244, USA

Received 24 March 1998; accepted 5 May 1998; Received by S. Yokoyama

Abstract

Proteasomes are large multisubunit particles that act as the proteolytic machinery for the ubiquitin-dependent proteolytic pathway. The core of this complex, the 20S proteasome, is made up of seven a-type and seven b-type subunits, arranged in an (a1–a7)(b1–b7)(b1–b7)(a1–a7) configuration. Previous work had shown that there exist alternative isoforms of the *Drosophila melanogaster* a4-type subunit, encoded by two distinct genes, *a4t1_dm* and *a4t2_dm*, and that these are expressed exclusively in the germline of the testes. We sought to investigate the evolutionary conservation of this phenomenon by screening for orthologs of the a4-type gene family in the distantly related *Drosophila* species, *D. virilis*. We isolated the *D. virilis* orthologs of the somatically expressed gene, *a4_dm*, and the testes-specific gene, *a4t2_dm*. We failed to find an ortholog of the other testes-specific gene, *a4t1_dm*. The *a4_dm* gene maps to the X chromosome at 12A-C, its product shares 90% amino acid identity with *a4_dm*, and it is expressed at high levels in both males and females. The other gene, *a4t_dm*, encodes a protein most similar to the testes-specific *a4t2_dm* proteasome subunit (59% a.a. identity), and it maps to position 27 on chromosome 2. The expression of the *a4t_dm* gene is testes-specific, like that of *a4t2_dm*. The existence of testes-specific a4-type subunits in two widely diverged subgenera of *Drosophila* suggests that these subunit isoforms have important functional roles in spermatogenesis. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Pros28.1; Spermatogenesis; *Drosophila virilis*; Gene family

1. Introduction

Most regulated proteolysis in eukaryotes occurs via the ubiquitin/proteasome pathway, whereby protein substrates are selected by the covalent attachment of multiple copies of the well-conserved polypeptide ubiquitin (Jentsch and Schlenker, 1995; Hochstrasser, 1996). The tagged protein is then recognized and bound to a large complex known as the 26S proteasome, which unfolds the protein, removes the ubiquitin tag, and feeds the polypeptide substrate into an inner degradative chamber where it is processively cleaved into small peptides (Coux et al., 1996). The proteasome's core is a 20S particle made up of four stacked rings of seven subunits each. The two outer rings are identical, and

consist of seven distinct, but related, a-type subunits. Similarly, the two identical inner rings are each made up of seven different b-type subunits. The arrangement of these subunits in the yeast, *Saccharomyces cerevisiae*, has recently been solved by X-ray crystallographic analysis (Groll et al., 1997), and the subunits have been renamed to conform to a standard nomenclature [see Zaiss and Belote (1997) for a list of the previously used synonyms for the yeast, fly and human subunits].

We are interested in the question of whether there are structurally different subpopulations of 20S proteasomes with functionally distinct properties. One example of proteasome heterogeneity is suggested by recent studies on *Drosophila melanogaster* proteasomes. Two alternative forms of one of the a-type subunits, a4 (also known as Pros28.1; Haass et al., 1990), encoded by two additional genes, are found to be expressed exclusively in the germline of the testes (*a4t1* and *a4t2*, Yuan et al., 1996). This finding is important because it represents a striking example of cell-type specific proteasome gene expression, and it suggests the possibility that 20S

* Corresponding author. Tel.: (315) 443 3695; Fax: (315) 443 2012; E-mail: jbelote@mailbox.syr.edu

Abbreviations: bp, base pair (s); kb, kilobase pair (s); MMLV, Moloney Murine Leukemia Virus; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase–polymerase chain reaction; S, Svedberg unit.

proteasomes with alternative subunit compositions exist in spermatogenic cells. Among the questions raised by this discovery are: (1) What is the functional significance of these testes-specific a-type subunits? (2) Do any of the other a- or b-type subunits also exist in testes-specific forms? (3) Are testes-specific a4 isoforms unique to *Drosophila melanogaster*, or is this an evolutionarily conserved feature? In this paper, we address this last question by searching for orthologs of these a4-related genes in the distantly related *Drosophila* species, *D. virilis*.

2. Materials and methods

2.1. General procedures

All standard techniques (e.g., phage and plasmid DNA purification, restriction enzyme digestion, gel electrophoresis, etc.), were done according to the protocols described in Sambrook et al. (1989). *D. virilis* flies were obtained from the Mid-America Stock Center (Bowling Green State University).

2.2. Genomic library screening

The *D. virilis* genomic library was an amplified library constructed by Ronald Blackman (Univ. of Illinois) using the bacteriophage λ EMBL3. The library was probed with a random primed 32 P-labeled 0.9-kb fragment from a cDNA clone, pGEM3/5D1, containing the entire coding region of the *D. melanogaster* gene a4 (Yuan et al., 1996). Reduced stringency hybridizations were done as described in O'Neil and Belote (1992). Hybridization was carried out overnight at 42°C in 5 × SSPE, 0.1% SDS, 1 × Denhardt's, 0.1 mg/ml sheared salmon sperm DNA, and 37% formamide. Filters were washed at 50°C in three changes of wash buffer (0.5 × SSPE, 0.1% SDS). Hybridizing phage were plaque purified, DNA extracted and the relevant restriction fragments subcloned into the plasmid vector pGEM3 (Promega, Madison, WI, USA) for subsequent analysis.

2.3. DNA sequencing

DNA sequencing was done by the method of Sanger et al. (1977). Primers for the sequencing reactions were either oligonucleotides corresponding to the T7 or SP6 promoters in pGEM3, or were gene-specific oligonucleotide primers obtained from Gibco/BRL (Gaithersburg, MD, USA). The sequences described here have been deposited into the GenBank database, accession numbers AF017649 and AF017650.

2.4. In situ hybridization to polytene chromosomes

Salivary gland squashes were prepared from late third instar *D. virilis* larvae, and in situ hybridization of biotinylated probes was done as described in Yuan et al. (1996). Chromosome localization was done using the cytological maps of *D. virilis* described in Gubenko and Evgen'ev (1984).

2.5. RNA analysis

For the Northern blot analysis, poly-A⁺ RNA was isolated from 0.5 g each of *D. virilis* adult males and females, and electrophoresed through a formaldehyde agarose gel as described in Sambrook et al. (1989). The RNA was transferred onto GeneScreen nylon membrane NEN (Boston, MA, USA) and serially probed with random primed labeled restriction fragments corresponding to the poorly conserved 3' portions of the a4_{dv} and a4t_{dv} coding regions. For the a4_{dv} gene, the 0.7-kb *EcoRI/HindIII* fragment containing most of exon 3 and 3' flanking sequence was used as the probe. For the a4t_{dv} gene, the 1.1-kb *EcoRV/SalI* fragment, containing the entire exon 3 and additional 3' sequences, was used as the probe (see Fig. 1). Between probings, the hybridizing counts were stripped from the filter by washing in boiling TE (10 mM Tris (pH 7.5), 1 mM EDTA). Filters were subjected to autoradiography for 6 days (for a4_{dv}) or 2 weeks (a4t_{dv}).

For the RT-PCR experiment, testes from ten adult males were hand-dissected in Ringer's solution, placed in a microfuge tube and quick frozen with liquid nitrogen. The carcasses from two of these flies were also frozen in a separate tube. Nucleic acid was prepared by grinding the tissue in RNA lysis buffer [7 M urea, 0.35 M NaCl, 0.1 M Tris-HCl (pH 8.0), 0.01 M EDTA,

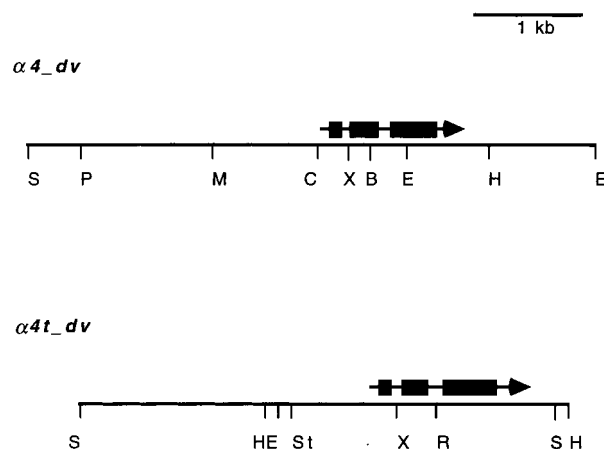


Fig. 1. Restriction maps of genomic regions containing the a4_{dv} and a4t_{dv} loci. The arrows represent the approximate positions and orientations of the genes, as inferred from sequence analysis. The solid boxes indicate the positions of the exons. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hind*III; M, *Mlu*I; P, *Pst*I; R, *Eco*RV; S, *Sal*I; St, *Stu*I; X, *Xba*I.

2% SDS], phenol extracted twice, and ethanol precipitated. After centrifugation, the pellet was dissolved in 50 ml DNase buffer [0.4 M Tris-HCl (pH 8.0), 0.1 M NaCl, 0.06 M MgCl₂, 0.1 M CaCl₂], and 1 unit of RNase-free DNase (Promega) was added and the samples incubated at 37°C for 1 h. The enzyme was then inactivated by heating to 65°C for 10 min, extracted twice with phenol:chloroform:isoamyl alcohol (24:24:1), and ethanol precipitated. The RNA was redissolved in 0.5 ml of diethylpyrocarbonate-treated water and the polyA⁺ RNA was then selected using an oligo-dT affinity column (Ambion, Inc. Austin, TX, USA) according to the manufacturer's instructions. The polyA⁺ RNA was eluted into 1 ml of buffer, ethanol precipitated and redissolved in 37 ml TE. This was used to synthesize cDNA using MMLV reverse transcriptase (Promega) as described by the supplier. For the PCR reaction, 2 ml of cDNA [or genomic DNA (1 ng) as a control] was used as template. The reaction cycles were as follows: 1 min 95°C, 1 min 55°C, 1 min 72°C for 35 cycles. The primers used were as follows: for the *a4_dv* gene, the 5' primer = 5'-ATAAATCGTGCCAGGTC-3' and the 3' primer = 5'-CCTTTTCAATTATGTC-3'; for the *a4t_dv* gene, the 5' primer = 5'-TGGAATACGCTCAAGAAG-3' and the 3' primer = 5'-GTGT-TATATACTCCACAGATGCGGGCCGATCAAA-3'.

2.6. Sequence analysis

The deduced amino acid sequences from *a4_dv* (GenBank accession number AF017649) and *a4t_dv* (AF017650) were aligned with those deduced from *a4_dm* (X62286), *a4t1_dm* (U46008) and *a4t2_dm* (U46009), using multiple alignment programs in CLUSTAL V (Higgins et al., 1992). The rat ortholog (D30804) was used as the outgroup to construct a rooted evolutionary tree for the *a4*-type isoforms. The topology and branch lengths of the phylogenetic tree were estimated by MEGA (Kumar et al., 1993) applying the neighbor-joining method on numbers of amino acid substitutions per site with Poisson correction (Nei, 1987). The reliability of the neighbor-joining tree was evaluated by bootstrap analysis with 1000 replications.

3. Results

3.1. Molecular cloning and mapping of *a4*-related genes from *D. virilis*

In order to isolate *D. virilis* genes corresponding to the *D. melanogaster a4* gene and its testes-specific forms, a *D. virilis* genomic library was screened under reduced stringency conditions using an *a4_dm* cDNA as a probe. Approximately 100 000 recombinant λ phage were screened resulting in the isolation of 12 positively hybrid-

izing plaques. After rescreening and plaque purification, the hybridizing fragments of these phage DNAs were subcloned into a plasmid vector and analyzed in more detail, and the results indicated that these phage represented sequences from two different loci. For reasons detailed below, these two loci will be referred to as *a4_dv* (contained on six of the recombinant phage) and *a4t_dv* (represented by six others). Fig. 1 shows the restriction maps of the genomic regions present on these two sets of overlapping recombinant phage.

The chromosomal positions of the *a4_dv* and *a4t_dv* loci were determined by situ hybridization of the cloned genes to *D. virilis* larval salivary gland chromosomes (Fig. 2). A recombinant phage carrying the *a4_dv* gene hybridized to the X chromosome [element A in Muller's terminology (Muller, 1940)] at cytological position 12A-C, and a *a4t_dv*-bearing recombinant phage hybridized to region 27 on chromosome 2 (Muller's element E). These chromosomes are thought to be homologous to

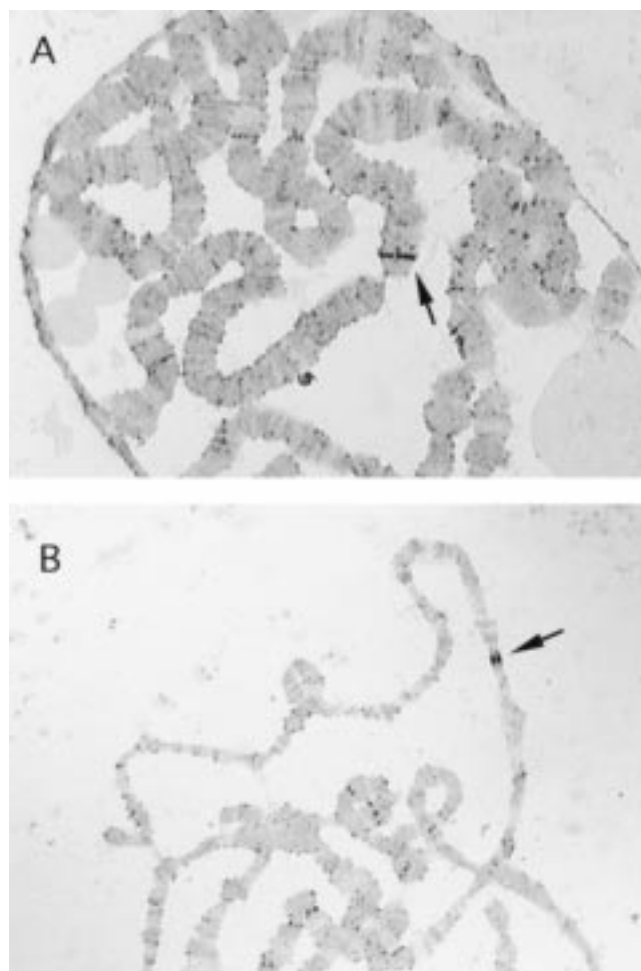


Fig. 2. In situ hybridization mapping of *a4*-related genes on *D. virilis* salivary gland chromosomes. (A) Localization of *a4_dv* to position 12A-C on the X chromosome. (B) Localization of *a4t_dv* to position 27 on chromosome 2. Photos taken under 400× magnification.

A

-421 ttgacaaatacaaaaaatcataaacgtattttaacggagcttacaatccggtttacatgattaacacggtaataatatta
 -341 tacaaccgtcacatagaaataacgataaaatcagacaatattttagacacaaacaaatgcacattatataatgggataacca
 -261 gggctggatatacaatctcttggctcatataaacgttggcaactagttttccatagcgataattatacaaaatttactaaaa
 -181 aaataccaaagatatacctggagtgctatcgataactatccggatagcaaacctcgcggccaccaggaatcccaaaa
 -101 aagcatcaatcctgcttgcacaaaaatctgcgactgcgatattttggatagcttaatatagctgcttcaacttaccgga
 - 21 aataataaaaactaaactaaagATGTCACCGTTACGATCGTGTACTATTTTTCGCCGATGTCGTTACTTCA
 M S S R Y D R A V T I F S P D G R L L Q 20

+ 60 AGTTGAATATGCTCAGGAAGCGGTTCGCAAGGTTCCACAGCGTgagtgccccatttggttatgtagtagagctca
 V E Y A Q E A V R K G S T A 34

+140 gatttaccatgtgtgcacactaatccatataatttataaagGTTGGTGTCCGTGGCGCAACTGTGTCTTACGCGT
 V G V R G G N C V V L G V 47

+220 GAGAAAAATCAGTAGCCAAAGCTCCAGGAAGCCGACAGTTCGTAATAATTTGCGTACTCGATCACGTTGTGATGCTTT
 E K K S V A K L Q E D R T V R K I C V L D H V V M A F 74

+300 TCGCGCCTGACTGATGCTCGCATCTTATAAATCGTCCAGGTCGAATGCCAGAGTCACTGCTGAATGTTGAGGATC
 A G L T D A R I L I N R A Q V E C Q S H R L N V E D 100

+380 CAGTGACCTTGAATATATAACAAGtaaaataaaatcgagatctcagccatatttgagtttctgtaattcaagaatgc
 P V T L E Y I T R 109

+460 ctaactcttgcagATATATTGCACAGCTGAAACAAAAGTACACTCAGAGCAATGGACGTCCTCCATTGGCATATCGTGC
 Y I A Q L K Q K Y T Q S N G R R P P G I S C 131

+540 CTTATTGGAGGCTTTGACGCTGATGGCTCACCTTATTATTTCAAAACCGAGCTTCTGGGATATCTTATGATGACAAAGC
 L I G G P D A D G S P Y L F Q T E P S G I F Y E Y K A 158

+620 CAATGCCACCGGCTTTCAGCAAAAGCCGTACGGGAATTTTGA AAAACAGTATCGCGAAGAGGAAGTTAGCACTGAAC
 N A T G C S A K T V R E F F E K Q Y R E E E V S T E 184

+700 GTGGTGTCTGAAAGTGGCCATTTCGGCCTGCTCGAAGTTGCCAATCCGGTCAACATAAATTTGGAGTGGCCATTATG
 R G A V K L A I R A L L E V A Q S G Q H N L E V A I M 211

+780 GAAAAATGCAAGCCTCGAAAAATGCTAGATCGCAAGTTATTTTGAATATTTTGGACATAATTTGAAAAGGAAAAGGAGGA
 E N G K P L K M L D R K V I L E Y L D I I E K E K E E 238

+860 GGAAC TTGAAAAGAAAGCAAGAAAAAGtaaatcccaatccatattttaaattagatggggaaatttgcactgcttattcttt
 E L E K K K Q K K * 247

+940 aaaaatcgtctttttaaatttcataattttataaattgaaagctgtacaatggaacaaaactttaaagatagtataaatac
 atggtcaaaaaaaatacaaaaaaacgagcctttaaattcaatgattacaattcaaatggctgggttctgtggtgg
 +1020 gtggaacaaaaataaaattaaactaaagcgctgtttacggtataagacaaaattatttggtaatttgg
 +1100

B

-386 gccttgctttaaataattggccatgcatttggcraataaacttctgttgcataaaagtcctattccagatcaaatggta
 -306 tccgatccgaatttcaaatatcgaaagctctcaacataaaaaacaatcaatagatgctgtgcaagcccaactctacatac
 -226 atacactgctcaacaaatttccgatcccaactaaatggcttggctgtttagttagtcaactgctccaggtgccagttttaa
 -146 atatatgtattactattaaactatacagcattcagtttctggctggaattaaggatggaagccataagaacatagca
 - 66 aacgggttggtttccatatacaaaatccgatctcaaatatgtataaattgctatttatagaATGTCCTCTTGA
 M S R P D 5

+ 15 TGAATCTGTAATAATATATATCCGATGGCCATTGTGCAAGTGAATACGCTCAAGAAGCACTGCCAAAAGTTCCACAG
 E S V N I Y S D G H L L Q V E Y A Q E A V R K G S T 31

+ 95 tggtaagtgcacaaatttggcagcatatataccatattctagaccagatcagccagctcagatagccagcaatcccg
 V G L R T K E C V V 41

+175 gcaagggatttccatattcttggctctgttctctttttttctcagctcagctGGGACTTCGCCACCAAGGAATGCGTGT
 V G L R T K E C V V 41

+255 GCTGGCCCTGGAGAAGAGCCATTGATTCCTTACAGATCGAGCCTACCTCGCGTAAGATTAAAGAAATAGATCAACATA
 L G V E K R A I D S L Q I E R T S R K I K K I D Q H 67

+335 TGGCAATGACATTTGCGGGCTTTCGGCCGATGCTCGCGTATTGGTGAAGTGCACCCAGCTGAGGCCAGAGCCACCGC
 N A M T F A G L T A D A R V L V S R A Q V E A Q S H R 93

+415 TTGAAC TTGATCGGCCCGCATCTTGTGGATATATAACAGCGTggggtgttccaccaagctgagtcagcgtgctcctgc
 L N F D R P A S V E Y I T R 108

+495 aactaacgcatcacagATATCTGGCCAGACTGAAAAGAGAGCTATCTCAGAGTGTGGCAGACCTCCGTTTGGCGTCC
 Y L A R L K Q T Y T Q S V A R R P P G V S 127

+575 TGTCTGATAGCGGCTTTCGTAAGATGGCAGGCGCGCTATTCAGACCCGATCCCTCGGCGATCTACTGAAATGGTC
 C L I G G F D E D G R P R L F Q T D P S G I Y Y E W S 154

+655 GGCCATACGACGGCCCGCTCGGCCAAAAGGTAACGATGACTGAGAAAGAACAGCTGCCATTTCAGGACCCAG
 A N T T G R L G Q T V N E Y L E K N T A A I S R T P 181

+735 ACGGCCCTCGGCCATGAAGCATGTGTGCTGCTGCTTGTTCACGGCCACATCGCTAGATCCAGCTTCTATTGAACTGGT
 D A A S A M K H V V R A L P T A T S L D P A S I E L A 208

+815 GTGCTCAGACTGCGACCCAAATCGAGATGGTCAAAACCGAAACGTTGGAATATCTGCTGGAGTCAATTAAGCAGGAGC
 V L R Y W Q P I E M V K P E T L E Y L L G V I K Q E A 235

+895 CGTGGAGAAGCGCTTACAAAAGCCCGGtaaacgctttagtccatagaactcgtttaaataatatttccgtaaacgctt
 V E E A L T K A P * 245

+975 tgaatgccatagaactcgttaataatatttctgctggtaaacgggtacacttgcataccgattgttttttggtttac
 +1055 aaatttggatattaaaatttccgaatatttgggtatagccacagcttatcatctaaaacgtgtatatacccttagagt
 +1135 taccaaatgtgcccagcttggatttccatatacaaaagatccgttgcactgagctgagcctgagcagctgcaagcttag
 +1215 tattctatagtgcaactaaatagcttggcgaatcagctagctatagctgttctcgtggaattgttatccgctcaaaa
 +1295 tccacacaaatcagagccggaagcaaaagtgaagcctggtgctcaatgagtgagtaactcaatcaatgctgct
 +1375 tggctcactgccccttccagtcgggaacctgtcgtccaagctgcatatgaatcggcaacgcccggggagaa

Fig. 3. DNA sequences of *D. virilis* genes: (A) *a4_dv*; and (B) *a4t_dv*. Upper-case letters represent coding sequences, and lower-case letters are non-coding regions. The numbers on the left represent nucleotide position relative to the beginning of the coding sequence. The numbers on the right represent the amino acid position. The underlined sequences indicate the positions of the primers used for the RT-PCR experiment (see Section 3.3).

the *D. melanogaster* X and 3R elements, respectively (Muller, 1940; Whiting et al., 1989).

3.2. DNA sequencing of the *D. virilis* a4-related genes

Subclones of the a4-hybridizing DNA fragments in these two genomic regions were sequenced, and the results are shown in Fig. 3. Examination of the sequences shows that both loci potentially encode proteasome subunits similar to those encoded by the *D. melanogaster* a4 gene family. Both *D. virilis* genes have two small introns, located in the exactly corresponding sites as the two introns in a4_dm and a4t2_dm (Frentzel et al., 1992; Yuan et al., 1996). Fig. 4 shows the alignment of the deduced amino acid sequences for all five *Drosophila* a4-type subunit genes. Comparison of the percent amino acid identity (Table 1) and the phylogeny (Fig. 5) derived from these data indicate that the a4_dv and a4t_dv genes are most closely related to a4_dm and a4t2_dm, respectively.

3.3. Sex-specific expression patterns of the *D. virilis* a4-related genes

The most striking feature of the *D. melanogaster* a4t1 and a4t2 genes is their testes-specific expression pattern.

Table 1
Pairwise comparisons of the *D. melanogaster* and *D. virilis* a4-type gene family

	% Amino acid identity			
	a4_dv	a4t1_dm	a4t2_dm	a4t_dv
a4_dm	90.4	73.9	59.6	51.2
a4_dv	—	73.1	58.4	52.4
a4t1_dm	—	—	53.4	49.2
a4t2_dm	—	—	—	58.8

Given the similarity between a4t_dv and a4t2_dm, it was expected that this gene would also be expressed in a testes-specific manner. As the initial step toward examining this, a Northern blot was done using polyA+ RNA extracted from adult *D. virilis* males and females. As can be seen in Fig. 6A, the a4_dv gene shows a similar expression pattern in the two sexes, as expected for the ortholog of the a4_dm gene, while the a4t_dv gene's RNA is male-limited. To confirm that this male-specific RNA is derived from testes-specific expression, RT-PCR methods were carried out to examine the tissue-specific nature of this RNA. In this experiment, cDNA was synthesized from polyA+ RNA isolated from either hand-dissected testes or the remaining carcass, and PCR-

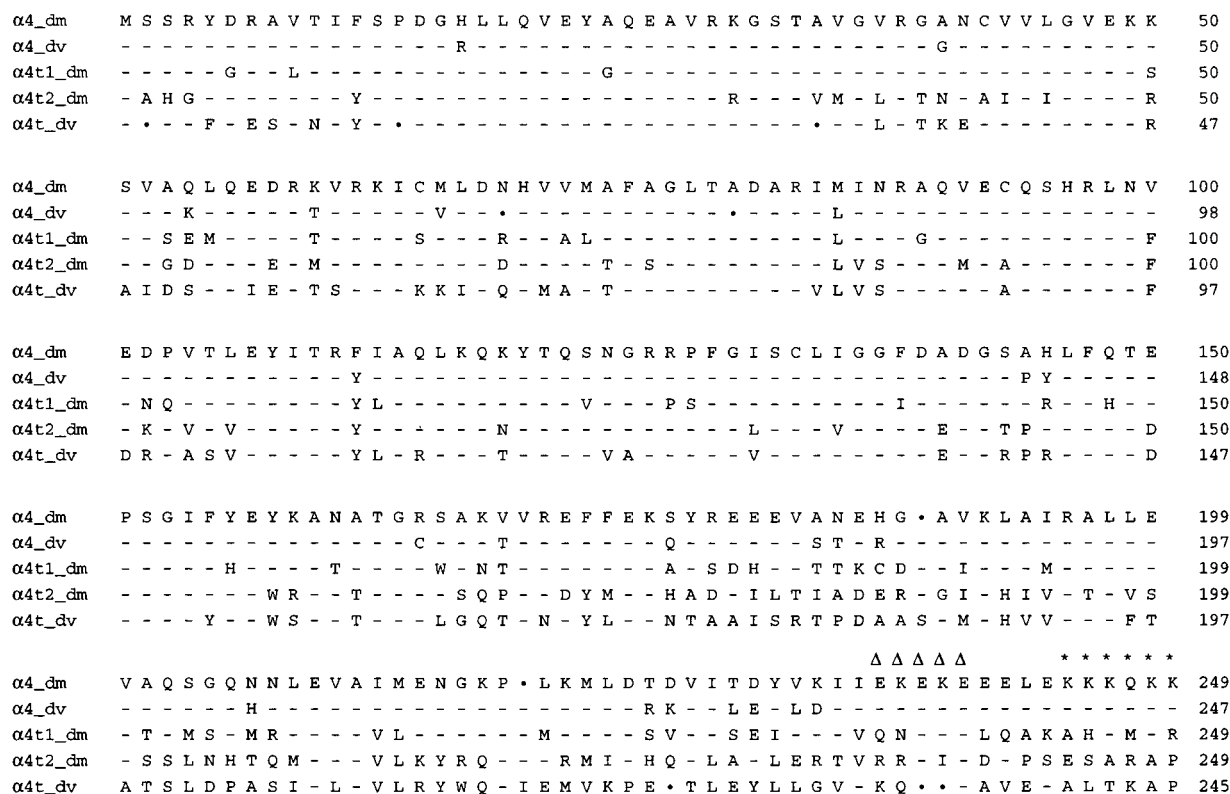


Fig. 4. Comparison of the inferred protein sequences of the a4-related proteasome subunits of *D. melanogaster* and *D. virilis*. The dashes indicate amino acids that are identical to those in the a4_dm subunit. The dots represent deleted amino acids. The numbers refer to the amino acid position at the end of each line. The triangles and asterisks highlight the KEKE motif and the putative nuclear localization signal, KKKQKK, respectively.

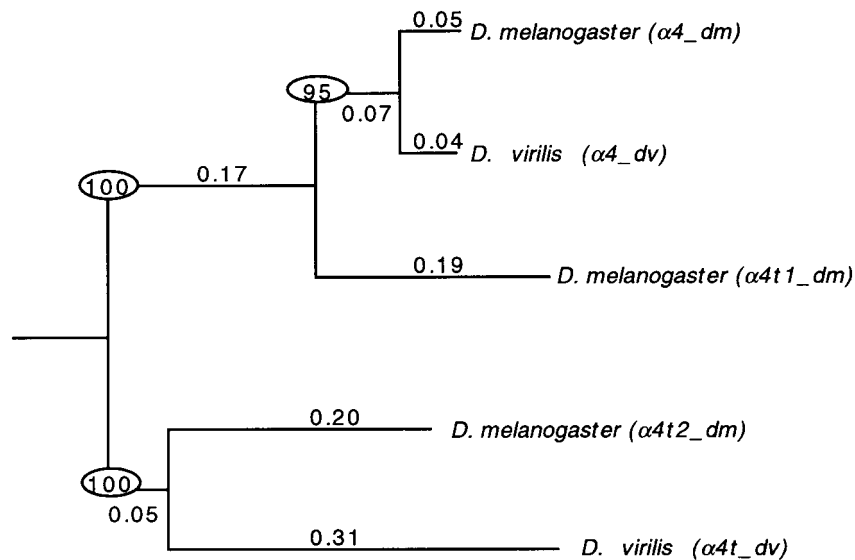


Fig. 5. Phylogeny of the *Drosophila* *a4*-type subunit genes based on deduced amino acid sequences. The numbers next to the different branches are branch lengths in amino acid substitutions per site.

amplified using primers specific for either *a4_dv* or *a4t_dv*. These primers were chosen to flank introns, so that the products derived from the mRNA could be distinguished from products that might be derived from contaminating genomic DNA. As seen in Fig. 6B, the *a4_dv* transcripts are found in both dissected testes and carcass, as would be expected for the somatically expressed form, while the *a4t_dv* transcript is detected only in the testes. RT-PCR using the *a4t_dv* primers consistently yields an additional prominent band of approximately 400 bp found in both testes and carcass. This probably represents amplification of contaminating genomic DNA, or possibly unspliced pre-mRNA, since it matches the estimated size of such a product.

Although all three genes *a4t1_dm*, *a4t2_dm* and *a4t_dv* exhibit similar testes-specific expression patterns, no putative transcriptional regulatory motifs which might underlie this expression pattern were identified by routine sequence inspection within 400 bp of the transcription start sites. Thus, either the testes-specific regulatory elements are complex, or different for these three genes, or they lie outside of the compared regions.

4. Discussion

The results presented above support the notion that the *a4_dv* gene is the ortholog of the *D. melanogaster* *a4_dm* and that *a4t_dv* is the ortholog of the testes-specific *a4t2_dm*. That is, the *a4_dv* gene shows highest sequence similarity (90.4% amino acid identity) to *a4_dm*, it is expressed at high levels in somatic tissues of both sexes, and, like *a4_dm*, it maps to the X chromosome. The *a4t_dv* gene shows highest sequence similarity to *a4t2_dm* (58.8% amino acid identity), its

intron–exon structure is identical to that of *a4t2_dm*, like *a4t2_dm*, its expression is testes-specific. Usually, orthologous genes are found on homologous chromosomes, but the *a4t_dv* gene maps to chromosome 2 (homologous to the *D. melanogaster* chromosome arm 3R) instead of chromosome 5 (Muller's element C), which is the homolog of the *D. melanogaster* chromosome arm 2R. However, there are other examples of *D. melanogaster* and *D. virilis* orthologs located on different chromosome elements [e.g., the histone gene cluster (cited in Gubenko and Evgen'ev, 1984); the maltose gene cluster (Vieira et al., 1997)].

Despite extensive searching, we failed to find any other *a4*-like genes. Given that the hybridization conditions used should allow the detection of an ortholog of the *a4t1_dm* if one exists, then our failure to find such a gene could possibly be the result of a non-representative genomic library, or it could be because such a gene does not exist in *D. virilis*. If there is only one testes-specific *a4*-like gene in *D. virilis*, then either (1) the *a4t1_dm* gene arose in the *D. melanogaster* lineage after its divergence from the *D. virilis* lineage [estimated to be about 40 million years ago (Russo et al., 1995)]; or (2) it arose before the *Drosophila* and *Sophophora* subgenera split, and was subsequently lost in the *D. virilis* lineage. The current data do not distinguish between these possibilities.

It is clear from an examination of Fig. 4 that the C-terminal portions of these proteins are much more divergent than the well-conserved N-terminal regions. A phylogenetic tree of *a4_dm*, *a4_dv*, *a4t1_dm*, *a4t2_dm* and *a4t_dv* was constructed by applying the neighbor-joining method of Nei (1987) to the numbers of amino acid substitutions per site, as determined using Poisson estimates. Relative rate tests (Nei, 1987) show that the

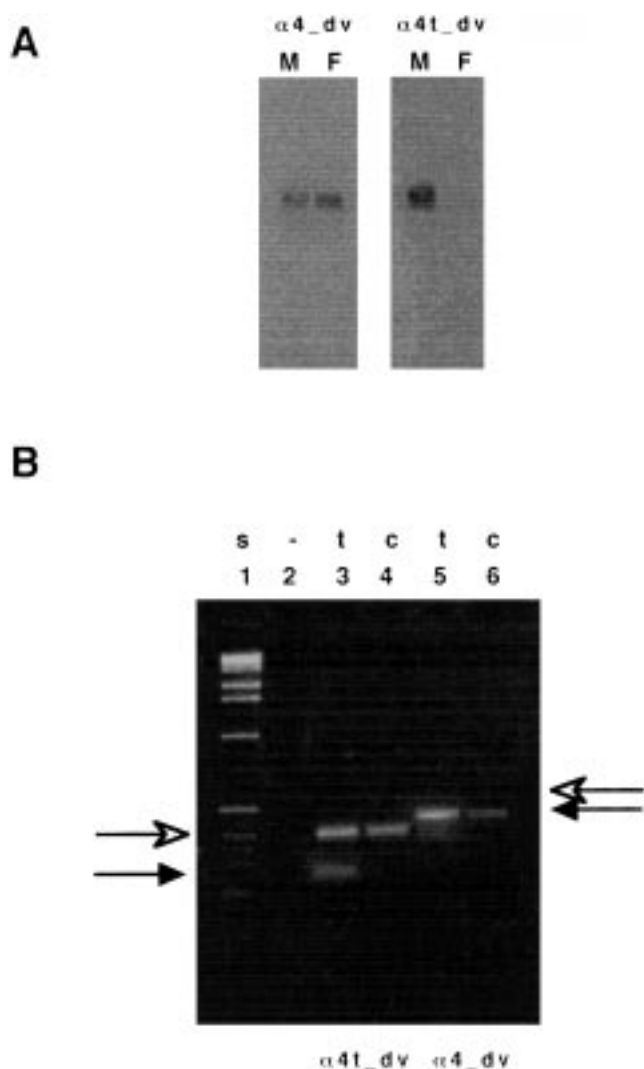


Fig. 6. RNA expression patterns of the *D. virilis* $a4$ -related genes. (A) Northern blot analysis of the expression of $a4_dv$ and $a4t_dv$ in adult males and females. The same filter was used for both probes, as described in Section 2.5. The left panel shows the results of hybridization with the $a4_dv$ probe and the right panel shows the results of hybridization with the $a4t_dv$ probe. M, male RNA; F, female RNA. (B) RT-PCR analysis of $a4_dv$ and $a4t_dv$. Lane 1, 1 kb ladder size markers (s); lane 2, blank; lane 3, testes (t); lane 4, carcass (c); lane 5, testes (t); lane 6, carcass (c). The samples in lanes 3 and 4 were amplified using primers specific for $a4t_dv$, and those in lanes 5 and 6 were amplified using primers specific for $a4_dv$. The filled arrows show the expected sizes of the RT-PCR products using primers specific for $a4t_dv$ (left, 267 bp) or $a4_dv$ (right, 455 bp). The open arrows indicate the expected sizes of products derived from contaminating genomic DNA (or unspliced pre-mRNA templates) for $a4t_dv$ (left, 399 bp) and $a4_dv$ (right, 523 bp, no band seen), respectively. The samples were run on a 2.5% agarose gel and stained with ethidium bromide.

$a4_dm$ $a4_dv$ genes are evolving at similar rates (ratio of branch lengths 0.94), as are $a4t2_dm$ and $a4t_dv$ when the testes-specific genes are compared (branch length ratio 0.82). In contrast, comparison of the three *D. melanogaster* $a4$ -type genes shows that the branch length to $a4t1_dm$ is 2-fold longer than that of $a4_dm$.

The branch lengths of all the testes-specific subunits are longer than their somatic counterparts. Thus, regardless of their source species or evolutionary relationship, the testes-specific genes appear to be evolving at a faster rate than the somatic $a4$ subunit genes.

One important question that remains concerns the functional significance of these alternative $a4$ -type subunits. Although the exact structure of the fly 20S proteasome is not known, if one assumes that its quaternary structure is similar to that of yeast, it would appear that the major differences between the testes-specific subunits and their somatic counterparts are situated on the outer surface of the particle, and near the regions that interact with the regulatory cap structures, such as the 19S or 11S complexes (Tanaka and Tsurumi, 1997; Kuehn and Dahlmann, 1997). However, direct evidence for specific interactions between the $a4$ -type subunits and particular regulatory subunits is lacking at this time.

Examination of the amino acid sequences does not reveal any conspicuous block of amino acids that is common to all of the testes-specific isoforms. However, it is seen that a putative nuclear localization signal [KKKQKK (Tanaka et al., 1990)], which is conserved among the $a4$ subunits of fly, human (Genbank accession number AF022815), rat (D30804), and chicken (U09226), is missing in all of the testes-specific forms. It has been shown that these sequences, as well as additional undefined sequences in the C-terminal 96 amino acids of the *D. melanogaster* $a4$ subunit, are important for nuclear localization (Kneuhl et al., 1996). Thus, it is possible that one functional consequence of the expression of the testes-specific subunits is an alteration in the subcellular localization of the proteasome during spermatogenesis. Also missing in the testes-specific forms is the KEKE motif, hypothesized to be important for protein-protein associations, perhaps with subunits in regulatory cap structures (Realini et al., 1994; Kuehn and Dahlmann, 1997). It is possible that a different 19S cap subunit binds to these testes-specific forms of the $a4$ -type subunits that does not require the KEKE motif for binding. This raises the question whether there are also testes-specific 19S cap subunits.

Whatever the exact functional consequences of testes-specific $a4$ -type subunits, the demonstration that this feature is conserved between *D. melanogaster* and *D. virilis* strongly suggests that alternative $a4$ -type proteasome subunits play a significant role in *Drosophila* spermatogenesis. It would be of interest to learn how widespread among other metazoans this phenomenon is, and what aspect of spermatogenesis is involved.

5. Conclusions

- (1) The *Drosophila virilis* genome contains at least two genes encoding $a4$ -type proteasome subunits.

- (2) One of these, *a4_dv*, is the ortholog of the *D. melanogaster* gene, *a4_dm*, and it maps to the X chromosome at 12A-C. It shares 90% amino acid identity with *a4_dm*, and is expressed at high levels in both males and females.
- (3) The other gene, *a4t_dv*, encodes a protein most similar to the testes-specific *a4t2_dm* proteasome subunit (59% a.a. identity), and it maps to position 27 on chromosome 2. The expression of the *a4t_dv* gene is, like that of *a4t2_dm*, testes-specific.
- (4) Intergenic sequence comparisons suggest that the testes-specific isoforms are rapidly evolving, particularly in their C-terminal regions. The functional significance of this is not known, but it could be related to differences in subcellular localization (e.g., nuclear vs cytosolic) and/or interactions with regulatory components of the proteasome.
- (5) The existence of testes-specific a4-type subunits in two subgenera of *Drosophila* that diverged over 40 million years ago strongly supports the notion that the expression of these subunit isoforms has important biological consequences.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. MCB-9506885. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Coux, O., Tanaka, K., Goldberg, A.L., 1996. Structure and functions of the 20S and 26S proteasomes. *Ann. Rev. Biochem.* 65, 801–847.
- Frentzel, S., Troxell, M., Haass, C., Presold-Hurt, B., Gladtscher, K.H., Kloetzel, P.-M., 1992. Molecular characterization of the genomic regions of the a-type subunit proteasome genes PROS-Dm28.1 and PROS-Dm35. *Eur. J. Biochem.* 205, 1043–1051.
- Groll, M., Ditzel, L., Lodwe, J., Stock, D., Botchler, M., Bartunik, H.D., Huber, R., 1997. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463–471.
- Gubenko, I.S., Evgen'ev, M.B., 1984. Cytological and linkage maps of *Drosophila virilis* chromosomes. *Genetica* 65, 127–139.
- Haass, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K., Kloetzel, P.-M., 1990. The PROS-28.1 gene is a member of the proteasome gene family. *Gene* 90, 235–241.
- Higgins, D.G., Bleasby, A.J., Fuchs, R., 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* 8, 189–191.
- Hochstrasser, M., 1996. Ubiquitin-dependent protein degradation. *Ann. Rev. Genet.* 30, 405–439.
- Jentsch, S., Schlenker, S., 1995. Selective protein degradation: a journey's end within the proteasome. *Cell* 82, 881–884.
- Kneuhl, C., Seelig, A., Brecht, B., Henklein, P., Kloetzel, P.-M., 1996. Functional analysis of eukaryotic 20S proteasome nuclear localization signal. *Exp. Cell Res.* 225, 67–74.
- Kuehn, L., Dahlmann, B., 1997. Structural and functional properties of proteasome activator PA28. *Mol. Biol. Rep.* 24, 89–93.
- Kumar, S., Tamura, K., Nei, M., 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, PA 16802.
- Muller, H.J., 1940. Bearings of the *Drosophila* Work on Systematics. In: Huxley, J. (Ed.), *The New Systematics*. Clarendon Press, Oxford, pp. 185–268.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- O'Neil, M.T., Belote, J.M., 1992. Interspecific comparison of the gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* 131, 113–128.
- Realini, C., Rogers, S.W., Rechsteiner, M., 1994. KEKE motifs. Proposed roles in protein-protein association and presentation of peptides by MHC class I receptors. *FEBS Lett.* 348, 109–113.
- Russo, C.A., Takezaki, N., Nei, M., 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* 12, 391–404.
- Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Tanaka, K., Tsurumi, C., 1997. The 26S proteasome: subunits and functions. *Mol. Biol. Rep.* 24, 3–11.
- Tanaka, K., Yoshimura, T., Tamura, T., Fujiwara, T., Kumatori, A., Ichihara, A., 1990. Possible mechanism of nuclear translocation of proteasomes. *FEBS Lett.* 271, 41–46.
- Vieira, C.P., Vieira, J., Hartl, D.L., 1997. The evolution of small gene clusters: evidence for an independent origin of the maltase gene cluster in *Drosophila virilis* and *Drosophila melanogaster*. *Mol. Biol. Evol.* 14, 985–993.
- Whiting Jr., J.H., Pliley, M.D., Farmer, J.L., Jeffrey, D.E., 1989. In situ hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. *Genetics* 122, 99–109.
- Yuan, X., Miller, M., Belote, J.M., 1996. Duplicated proteasome subunit genes in *Drosophila melanogaster* encoding testes-specific isoforms. *Genetics* 144, 147–157.
- Zaiss, D., Belote, J.M., 1997. Molecular cloning of the *Drosophila melanogaster* *a5_dm* encoding a 20S proteasome a-type subunit. *Gene* 201, 99–105.