Structure and tissue-specific expression of the *Drosophila melanogaster* organellar-type Ca^{2+}-ATPase gene

Attila MAGYAR,* Éva BAKOS and András VÁRADI†
Institute of Enzymology, Biological Research Centre, Hungarian Academy of Sciences, P.O. Box 7, Budapest, 1502 Hungary

A 14 kb genomic clone covering the organellar-type Ca^{2+}-ATPase gene of *Drosophila melangaster* has been isolated and characterized. The sequence of a 7132 bp region extending from 1.1 kb 5′ upstream of the initiation ATG codon over the polyadenylation signal at the 3′ end has been determined. The gene consists of nine exons including one with an exceptional size of 2172 bp representing 72% of the protein coding region. Introns are relatively small (< 100 bp) except for the 3′ intron which has a size of 2239 bp, an exceptionally large size among *Drosophila* introns. Five of the introns are in the same positions in *Drosophila*, *Artemia* and rabbit SERCA1 Ca^{2+}-ATPase genes. There is only one organellar-type Ca^{2+}-ATPase gene in the *Drosophila* genome, as was shown by Southern-blot analysis [Váradi, Gilmore-Hebert and Benz (1989) FEBS Lett. **258**, 203–207] and by chromosomal localization [Magyar and Váradi (1990) Biochem. Biophys. Res. Commun. **173**, 872–877]. Primer extension and S1-nuclease assays revealed a potential transcription initiation site 876 bp upstream of the translation initiation ATG with a TATA-box 23 bp upstream of this site. Analysis of the 5′ region of the *Drosophila* organellar-type Ca^{2+}-ATPase gene suggests the presence of potential recognition sequences of various muscle-specific transcription factors and shows a region with remarkable similarity to that in the rabbit SERCA2 gene. The tissue distribution of expression of the organellar-type Ca^{2+}-ATPase gene has been studied by *in situ* RNA–RNA hybridization on microscopic sections. A low mRNA abundance can be detected in each tissue of adult flies, suggesting a housekeeping function for the gene. On the other hand a pronounced tissue specificity of expression has also been found as the organellar-type Ca^{2+}-ATPase is expressed at a very high level in cell bodies of the central nervous system and in various muscles.

INTRODUCTION

Intracellular organelles store Ca^{2+} and the cytoplasmic concentration of Ca^{2+} is partly controlled by its accumulation or release from these organelles. Organellar-type Ca^{2+}-ATPases pump Ca^{2+} back into the storage organelles after release. These enzymes belong to the family of P-type ion-motive ATPases together with, for example, the Na^+,K^+-ATPase, H^+-ATPase and the plasma membrane Ca^{2+}-ATPase. These enzymes share common molecular architecture, basic mechanistic features of ion transport and show sequence homology. In mammals three different genes encode the organellar-type (sarco/endoplasmic reticulum-type) Ca^{2+}-ATPases, SERCA1, SERCA2 and SERCA3 [1–3]. (The nomenclature for the sarco/endoplasmic reticulum-type Ca^{2+}-ATPases proposed by Burk et al. [3] is followed.) Both the fast-twitch skeletal muscle form (SERCA1) and the cardiac (slow-twitch) skeletal muscle form (SERCA2) appear in two alternatively spliced variants [4–6]. A similar type of alternative splicing of the chicken SERCA2 gene has also been revealed [7].

Homologous gene products with approx. 70% amino acid sequence identity with their vertebrate sarco/endoplasmic reticulum-type Ca^{2+}-ATPases counterparts were described in invertebrates, namely in *Artemia franciscana* [8] and in *Drosophila melanogaster* [9]. In contrast to the vertebrate SERCA-gene families, there is only one SERCA gene in both invertebrates studied [9–11]. Alternative splicing events similar to those described for vertebrate SERCA2 genes have been described for the *Artemia franciscana* sarco/endoplasmic reticulum-type Ca^{2+}-ATPase gene [12]. The complete structure of SERCA genes are somewhat less known than that of their products. The exon/intron organization of the rabbit SERCA1 gene [13] and partial characterization of the rabbit, pig and human SERCA2 genes [14–16] have been published. Recently the structure of the *Artemia franciscana* SERCA gene became known [11]. The gene structure of two more distant homologues, the organellar-type Ca^{2+}-ATPase gene of *Plasmodium* [17] and that of tomato [18] have also been described recently.

Previously we have cloned and sequenced the cDNA of the organellar-type Ca^{2+}-ATPase of *Drosophila melanogaster*, mapped the gene to the end of the right arm of chromosome 2 at band 60A-B, demonstrated that its expression is developmentally regulated and found that there is only one gene in the fly genome [9,10].

In order to understand better the evolution of this gene family, as well as the mechanism of regulation of the Ca^{2+}-ATPase gene expression in *Drosophila melanogaster*, we have embarked upon a study to determine the structure of the gene on the DNA-sequence level as well as its expression in different tissues of adult fly.

EXPERIMENTAL

Library screening and DNA-sequencing

The genomic library of adult *Drosophila melanogaster* constructed in EMBL4 phage was provided by Dr. A. Udvardy

Abbreviations used: DTT, dithiothreitol; SSC, 0.15 M NaCl/0.015 M sodium citrate.

* Present address: Department of Biochemical and Biophysical Sciences, University of Houston, Houston, TX 77204-5934, U.S.A.
† To whom correspondence should be addressed.

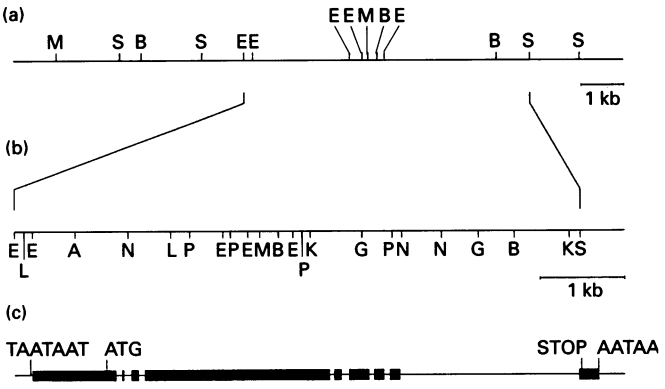

Figure 1 Structure of the *Drosophila* organellar-type Ca^{2+}-ATPase gene

(**a**) The restriction map of the 14 kb genomic clone, DgCa14; (**b**) represents the 7.2 kb region whose DNA sequence has been determined and (**c**) is a schematic representation of the exon/intron structure of the gene. Solid bars represent exonic while narrow lines represent intronic regions. Restriction enzymes used for mapping: A, *Apa*I; B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; K, *Kpn*I; L, *Cla*I; M, *Sma*I; N, *Nsi*I; P, *Pst*I; S, *Sal*I.

(Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Hungary). For screening under high-stringency conditions [19] two cDNA probes were used; the 5′ probe was an *Eco*RI fragment corresponding to cDNA region 33 to 1145, while the other probe was derived from the most 3′ end of the coding region (1968 to 3206 bp) by *Eco*RI digestion and contained some of the 3′ untranslated region of the cDNA [9]. By screening approx. 2×10^5 independent plaques several positive recombinant phages giving positive hybridization signals with both of the probes have been identified. The largest one, DgCa14, was further analysed with restriction mapping and with Southern blotting (Figure 1). This is a 14 kb genomic clone containing all of the exon sequences, an approx. 5 kb 5′ non-transcribed region and an approx 2 kb 3′ non-transcribed region. The DNA sequences of a 7.2 kb region of DgCa14 from the initiation ATG codon over the polyadenylation signal at the 3′ end as well as a 1.1 kb 5′ region upstream of the initiation ATG codon have been determined (Figure 1).

Primer extension

Total RNA was isolated from adult flies using the guanidinium thiocyanate method [20]. Two antisense oligonucleotide primers, BE05: 5′-CTTGATTCGTTATCCTCTTCGTTCAGTGGCCT and BE03: 5′-ATTTACTGCCTATCGCCGTTCTGCCTCA corresponding to nucleotides −32 to −1 bp and −839 to −812 bp were 5′ end-labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase to the specific activity of 10^8 c.p.m./mg. Primer extension analysis was performed as described [19]. In summary, in each experiment 10^6 c.p.m. of primer was hybridized with 35 μg of total RNA overnight at 32 °C in hybridization buffer (40 mM Pipes, 1 mM EDTA, 0.4 M NaCl, 80% formamide). Reverse transcription reaction proceeded with 200 units of Moloney murine leukaemia virus reverse transcriptase (Superscript II; Gibco BRL) at 42 °C for 2 h in a reaction buffer consisting of 75 mM KCl, 50 mM Tris/HCl, pH 8.3, 3 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 0.3 mM EDTA, 1 mM dNTP and 20 units of RNasin. Following the reaction the samples were treated as described previously [19] and analysed on 4.5% or 6% polyacrylamide/7 M urea sequencing gels along with the products of DNA sequencing reactions using the appropriate 5′ genomic M13 subclone and the same primers.

S1-nuclease protection

Two different experimental approaches have been used: (i) hybridization of RNA to a double-stranded ^{32}P-labelled DNA probe as described [19]; and hybridization of RNA to a single-stranded probe generated by a solid-phase synthesis technique (see below).

The double-stranded probe was generated by an M13 single-stranded DNA clone as template carrying a genomic insert of −1082 to 1379 and BE05 oligonucleotide as primer (see Figure 4). Restriction enzyme *Hin*dIII (10 units) was present during second-strand synthesis to prevent chain-extension over the polycloning site. Hybridization to 35 μg of total RNA prepared from adult flies was carried out at 52 °C; S1-nuclease digestion was performed at 37 °C according to the standard procedure [19].

The single-stranded antisense DNA probe corresponding to nucleotides −812 to −1083 bp was prepared by a solid-phase technique as follows: the *Eco*RI genomic fragment (−1082 to 2206 bp) cloned into pBluescript SK$^+$ (Stratagene) was used to prepare the probe. This construct was digested with *Nsi*I (at 264 bp) and the 3′ ends were biotinylated using 45 units of Terminal Transferase (Boehringer) and 0.1 mM biotin-16-dUTP (Boehringer) in a reaction buffer of 200 mM potassium cacodylate, 25 mM Tris/HCl, pH 6.6, 25 mg/ml BSA, 2 mM $CoCl_2$ at 37 °C for 1 h. After ethanol precipitation the −1083 to 264 bp insert was removed by *Kpn*I digestion and gel-purified. The biotinylated double-stranded fragment was attached to streptavidin-coated magnetic beads (Dynebeads M-280 Streptavidin; Dynal), then denatured with 0.1 M NaOH and the strands were separated magnetically as described [21]. The antisense DNA probe was generated from the immobilized single-stranded DNA template. The immobilized DNA together with 2 pmol of BE03 primer in 10 μl of annealing buffer (40 mM Tris/HCl, pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl) was heated to 65 °C and allowed to cool to room temperature. [α-^{32}P]dATP (50 μCi) and 1 μl aliquots of dCTP, dGTP, dTTP (500 μM each) were added to the DNA/primer annealing mixture. The extension step was performed at 37 °C for 30 min using T4 DNA polymerase (Sequenase II; USB), then 1 μl of 500 mM chase dATP was added and incubation was continued for an additional 10 min. The synthesized radiolabelled probe was eluted with 0.1 M NaOH and neutralized. Probe (5×10^5 c.p.m.) was hybridized with 35 μg of total RNA under the same conditions as described for primer extension. Then S1-nuclease digestion was performed as described [19] with 100 units of S1 nuclease at 37 °C for 1 h. The protection products were analysed on a 6% polyacrylamide/7 M urea sequencing gel alongside a sequence ladder generated with BE03 primer and a 5′ genomic M13 subclone as template.

In situ hybridization of microsections

Frozen sections (8 μm) were placed on poly(lysine)-coated slides and dried for 1 h at room temperature. The sections were fixed in 4% paraformaldehyde in PBS for 30 min at room temperature. After several washes in PBS the sections were dehydrated with rinsing in ethanol solutions by stepwise increase of the ethanol concentration from 30% to 100%. Prior to hybridization the sections were rehydrated in ethanol solutions of decreasing concentrations, 95% to 30%, incubated in 2 × SSC (SSC: 0.15 M NaCl/0.015 M sodium citrate) at 70 °C for 30 min and then treated with Pronase (0.02 mg/ml) for 5 min at room temperature. After several washes in PBS containing 0.2%

glycine the paraformaldehyde fixation step was repeated. Sections on ten microscope slides were incubated in 400 ml of 0.1 M triethanolamine hydrochloride, pH 8.0, containing 1 ml of acetic anhydride for 10 min at room temperature. This solution was freshly made and used immediately. After rinsing in PBS the sections were dehydrated as above. The hybridization mixture contained 50% formamide, 0.5 M NaCl, 10 mM Tris/HCl, pH 8.0, 20 mM DTT, 1 mM EDTA, 0.5 mg/ml yeast tRNA, 10% (w/v) dextran sulphate and 1 × Denhardt's solution [19]. RNA probes labelled with ^{35}S were generated with the Riboprobe (Stratagene) technique according to the manufacturer's suggestions using the full-size cDNA [9] cloned into Bluescript SK^+ as template. Both sense and antisense RNA probes were prepared and used. Sections on one slide were overlaid with 35 μl of hybridization mixture containing 0.25 mg/ml radiolabelled RNA probe, covered with a 22 mm coverslip and incubated in a well-sealed moisture chamber at 50 °C overnight. High stringency washing of the sections was performed in 2 × SSC/0.2% β-mercaptoethanol/50% formamide at 58 °C for 30 min. Then, after a brief wash in 2 × SSC, the section were subjected to RNase A digestion in 200 ml of DNase-free RNase A solution (20 ng/ml in 0.5 M NaCl, 0.1 M Tris/HCl, pH 8.0) at 37 °C for 30 min. A second wash in 2 × SSC/0.2% β-mercaptoethanol/50% formamide at 50 °C for 30 min was applied which was followed by a third washing step in the above mixture but omitting formamide at 50 °C for 30 min. Sections were dehydrated in ethanol as described above. The slides were dipped into autoradiography emulsion mixture (Kodak Type NTB2) and exposed at 4 °C for 2–4 days. After development the sections were stained with Giemsa as well.

RESULTS AND DISCUSSION

Structure of the *Drosophila* organellar-type Ca^{2+}-ATPase gene

Figure 1 shows the structure of the organellar-type Ca^{2+}-ATPase gene and the restriction map of 14 kb genomic DNA covering the gene. In order to obtain genomic clones a *Drosophila* genomic library was screened with ^{32}P-labelled cDNA probes under high-stringency conditions. Two cDNA probes were used; the 5′ probe was an *Eco*RI fragment corresponding to cDNA region −33 to 1145; while the other probe was derived from the most 3′ end of the coding region (1968 to 3206 bp) by *Eco*RI digestion and contained some of the 3′ untranslated region of the cDNA [9]. By screening approx. 2×10^5 independent plaques, several positive recombinant phages were identified. The largest one, DgCa14, hybridizing with both the 5′ and the 3′ cDNA probes, was further analysed with restriction mapping and with Southern blotting. The DNA sequence of a 7.1 kb region (7132 bp) starting 1.1 kb upstream of the translation initiation ATG, containing all the exon and intron sequences and extending over the poly-adenylation signal, has been determined and deposited in the EMBL databank (accession number: X84681).

Comparison of the published cDNA sequence [9] and the genomic sequence revealed two differences: a G → A in position 1314 and a T → C in position 2520 (cDNA numbering); both base variations affect third codon positions and mean silent mutations.

The *Drosophila* organellar-type Ca^{2+}-ATPase gene consists of nine exons with very wide size distribution (range 18–2172 bp); 72% of the protein is encoded by a large exon (Figure 1). The first exon contains the whole 5′ non-translated region (876 bp) and 118 bp of the coding sequence, while the 3′ exon contains 23 bp translated and 210 bp 3′ non-translated regions. The exon/intron structure of the *Drosophila* organellar-type Ca^{2+}-ATPase gene does not reflect the organization of the proposed functional domains of the protein. The same observation can be made by analysing the rabbit SERCA1 gene [13] or the *Artemia* Ca^{2+}-ATPase gene [11].

Figure 2 Comparison of the exon/intron structures of various organellar-type Ca^{2+}-ATPase genes

Vertical arrows show identical intron positions of genes next to each other in the Figure. Data are taken from the literature: rabbit SERCA1 gene [13], *Artemia franciscana* gene [11], tomato Ca^{2+}-ATPase gene [18] and *Plasmodium yoelli* Ca^{2+}-ATPase gene [17] while the structure of the *Drosophila melanogaster* Ca^{2+}-ATPase gene is from this study.

Table 1 Exon–intron boundaries in the *Drosophila* organellar-type Ca-ATPase gene

	Exon	position	Intron	size (bp)	Exon
IVS-1	CCC AAT G	118	/gtgaga-(	66)-	ttatcgttacag/AG TTG CCG
	P N				E L P
IVS-2	GAG GAA G	136	/gtgcgg-(	78)-	tgtattccacag/GA AAG AGC
	E E				G K S
IVS-3	TCA TTT	219	/gtaagt-(	67)-	tgcaatccccag/GTT CTC
	S F				V L
IVS-4	AAC TTG	2392	/gtactc-(	58)-	acttggtcacag/GTT ACT
	N L				V T
IVS-5	TTC TTC CG	2507	/gtgagt-(	63)-	tcgaacttttag/T TAC ATG
	F F R				Y M
IVS-6	ATG AAC AG	2744	/gtgagt-(	64)-	ttgtttaaacag/C TTG TCC
	M N S				L S
IVS-7	CTC TCC	2862	/gtaagt-(	68)-	ttttacccacag/ACC GTC
	L S				T V
IVS-8	GCA GAT G	2980	/gtgaga-(2239)-		ccgaaattgtag/TT CCT GAT
	A D				V P D

Although the five Ca^{2+}-ATPase transcripts shown in Figure 2 are approx. the same size, the number of exons of these gene products are very different: the rabbit SERCA1 gene encodes 22 exons, the *Artemia* SERCA gene contains 17, the *Drosophila*

The exon/intron structure of the *Drosophila* organellar-type Ca^{2+}-ATPase gene was aligned with the exon/intron structure of each Ca^{2+}-ATPase gene whose structure is known: the *Artemia* SERCA gene [11], rabbit SERCA1 gene [13], tomato organellar-type Ca^{2+}-ATPase gene [18] and *Plasmodium yoelli* Ca^{2+}-ATPase gene [17] in Figure 2.

```
GM-PLSSYFVDAWGLVLAWALFFGVIFYSPLstop  Artemia isoform B

GESPIYKMH----GIVLMWAVFFGLLYAMMLstop  Drosophila IVS8

*  *^          *^***^***^***^^^^  *
```

Figure 3 A putative alternatively spliced region in the 3′ intron

The amino acid sequence obtained by in-frame translation of the 5′ end of the 3′ intron is compared with the corresponding region of the *Artemia franciscana* Ca^{2+}-ATPase gene. This region is present in the *Artemia* B isoform but absent from isoform A [12]. Identical residues are indicated by asterisks, similar residues by ∧.

gene nine, the tomato gene seven and the *Plasmodium yoelli* gene consist of only three exons. Five of the eight introns interrupting the *Drosophila* transcript share identical positions with those of the *Artemia* gene as well as with the corresponding introns of the rabbit SERCA1 gene. Interestingly, IVS2 of the *Drosophila* gene is preserved in the rabbit SERCA1 gene too, but is absent from the *Artemia* gene. This intron is also present in the rabbit SERCA2 gene whose partial structure has been published [14]. No common intron position of the *Drosophila* gene and either the tomato or the *Plasmodium* gene can be detected. One of the introns of the tomato gene appears to be in an identical position with IVS13 of the rabbit SERCA1 gene, but this intron is absent from the *Drosophila* gene as well as from the *Artemia* gene. The most plausible explanation for the fate of this intron during the evolution of the organellar-type Ca^{2+}-ATPase genes is that this intron was present in an early ancestor 'Ca^{2+}-ATPase' gene which existed before the split of the animal and the plant kingdoms. This intron was selectively deleted from the *Drosophila* and *Artemia* genes after the split of vertebrate and invertebrates.

The five introns with identical positions in the *Drosophila*, *Artemia* and rabbit SERCA1 genes were most probably present in the common ancestor of the invertebrate and vertebrate Ca^{2+}-ATPase genes. The same is true for those seven introns whose positions are preserved in the *Artemia* and rabbit genes, but were absent from the *Drosophila* gene. These introns had been selectively lost from the *Drosophila* gene during its evolution. On the same line of reasoning IVS2, which can be found both in the *Drosophila* and in the rabbit SERCA genes, but is missing from the *Artemia* gene, had probably existed in the common ancestor of the invertebrate and vertebrate Ca^{2+}-ATPase genes and had been selectively lost from the *Artemia* gene.

The DNA sequences at the exon/intron boundaries, i.e. the splice donor and splice acceptor sites, are similar to the consensus sequence motifs found in *Drosophila* [22] as is shown in Table 1. The intron phase along the gene sequence does not show any

```
 (MCBF)                          (MCBF)       (chk. myoD)        (chk. myoD)
 CATTCCT                         CATTCCT      CCTCAGCTGTC         GGAACAGGTG     -1002
GAATTCCTTCGAAACTTTGATTAGTTTTAAATTCCTATATCACCTCAGCTATCGGTATTACTGAACACGGAAACTCAACA
                                                    -960
GTTAGCCATCGCAGTTGAGATTCTCAGTTATCGATAACGCCGGCTATCGCAGTTTAGTAGCACAAATTGAACGCGTCAGT
                          TATA-box          -880      *S1    *pr.ext.
AGAAGCTGAGCTAAAAGGTGGGATAATTATCTAATAATTGCCAGGACTGAGAATTCTTAAAAGTTGGAGAAAGAGGCAGC
                                                    -800                          <--
TGTGCACAAATAACGTAACTCGGACGATATACGTTTTCAGTCAGCCCTGTCTTGTGCGAATAATGTCGTGTCATAGTGAG
-------primer-BE03---                              -720
GCAGAACGGCGATAGGCAGTAAATCGCGGCTTGGTACTTAGTGCAATAGTTATCAGCACACATATTCAGAAAAAAGCGCC
                                                    -640
ATGGGTTATATTATATAGAGAGTCAGTGGAAAAAAGTACTTAACACACGCAGTGCGTCGTTTACGCAGGTTAACGTAGGA
                                                    -560
GCAGAGCACCCGTATTACGGACCAGATCCCCCAATCCCCCGCAGAAACTGAGAATAGAAAAACGAAAACTGCGTCTGTTG
                                                    -480
TGCGAAGTGACACGTGTGTGAATCTCATAAGCGGAGCGATTTGGCCAGGGTAACAGCCCATCATAGTAATGCAATATTCC
                                                    -400^^^^^^^^^^^^^^^
AGCATATCTCCGACCCGATCCGCACAATCCGATCCTAAGTTGGCGCGATAACTGCGCGACTTTAGTGGCCAGTCCGGGCC
                                                    -320
CGATCAGAAGTGCCTGAAATCGGGGAATTAGCATGTAGCACCTGTGTATGTGCGTATGTTGGAAAGTACCCGTGAGCCGA
                                                    -240
GTGCCGTGCACATGCACATGTACGACCACTTATTCAGAACCTAATGAGGTAGTTTTAACTGCAAACGTACATGCCAGTAC
                                                    -160                          --CArG-
AGATGTATGTGCATACATACGTGCATACACACTTGCTATTCCAGTCAGTCGTTGAATAGAGCCCAGAAAACGGCAAAATA
--                                                  -80
GCTAAATAGCCCTGCCAGAGAATTAAAATGAGCCCCTTGCCTTGTGTCTACTTTCAAATCCCCACCTGAATCGCCCCGGA
                                                      -1 1
ATCTGGACACAGGCCACTGAACGAAGAGGATAACGAATCAAG ATG
        <--------------primer  BE05---  Met
```

Figure 4 Sequence of the 5′ exon and 5′ untranscribed region of the *Drosophila* organellar-type Ca^{2+}-ATPase gene

Numbering starts at the translation initiation ATG; primers used for primer extension experiments are indicated below the sequence as dashed arrows together with the primer's name; the results of S1-nuclease mapping and that of the primer extension experiment are indicated by asterisks followed by S1, and by pr.ext., respectively; the TATA-box as well as various putative transcription factor recognition motifs are underlined and indicated above the sequence. *MCBF* refers to M-CAT binding factor [24]; *chk. myoD* to the MyoD1 site in the chicken acetylcholine receptor α subunit gene [25]; *CArG* to the CA-rich G-type muscle-specific enhancer recognition sequence [26]. The MyoD family consensus sequences (CANNTG) are double underlined. Region showing similarity to the 5′ end of the rabbit SERCA2 gene is indicated by ∧∧∧∧∧∧∧ below the sequence.

regularity; three phase 1 introns, two phase 2 introns and three phase 0 introns are present. The size distribution of the eight introns is in the range of 76–2257 bp. Seven of the eight introns are relatively small and fell into a similar size range, 76–96 bp, while one of them, IVS-8, has an extension of 2257 bp, which is an unusually large size among *Drosophila* introns. As a rule *Drosophila* genes (and those of eukaryotes with relatively short generation time) have relatively few introns and their introns are small, i.e. the size may represent a minimal structure capable for splicing. Introns with, for example, regulatory functions, however, can be much larger [23]. All introns of the *Drosophila* Ca^{2+}-ATPase gene fall into the first category with the exception of IVS-8 (Table 1). This 3′ intron is present in the very same location in the rabbit SERCA1, in the rabbit SERCA2 and in the *Artemia franciscana* Ca^{2+}-ATPase genes (see [13], [14] and [11] respectively); and these intron sequences are involved in alternative splicing events. This may be true for the *Drosophila* IVS-8 as well since the hypothetical amino acid sequence encoded by the 5′ end of this intron is rather similar to the C-terminal sequence of the alternatively spliced *Artemia* Ca^{2+}-ATPase isoform B (Figure 3). Furthermore, a potential polyadenylation signal (attaaa) is located in IVS8 at position 1277 downstream of the 5′ splice junction site. Organellar-type Ca^{2+}-ATPase genes giving SERCA2b-like alternative transcripts (the mammalian SERCA2 gene and the *Artemia* gene) also contain a polyadenylation site within their 3′ intron (before the last exon).

The occurrence of this type of alternative splicing event in the *Drosophila* Ca^{2+}-ATPase pre-mRNA remains to be established.

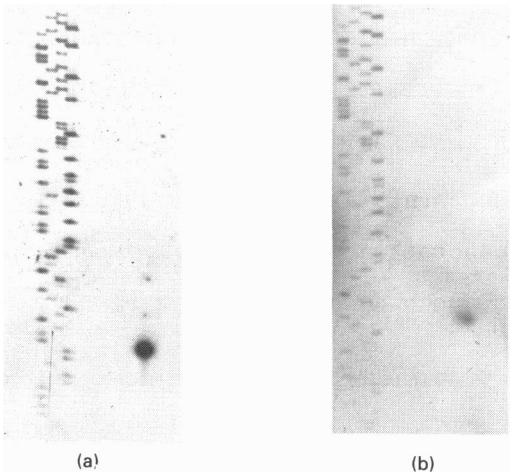

Figure 5 Identification of the transcription initiation site in the *Drosophila* organellar-type Ca^{2+}-ATPase gene

(**a**) Result of a primer extension experiment: end-labelled primer BE03 (see Figure 4) and RNA purified from adult flies have been used. (**b**) Result of an S1-nuclease protection assay. Antisense radiolabelled single-stranded DNA probe was generated by a solid-phase technique as described in the Experimental section. The probe, complementary to the region extending from positions −818 to −1083, had been hybridized with RNA purified from adult flies. A sequence ladder obtained with primer BE03 and an M13 genomic clone containing the 5′ region of the gene has been used as size marker on both (**a**) and (**b**).

Analysis of the 5′ end of the Ca^{2+}-ATPase transcript

The search for probable transcription initiation site(s) has included primer extension and S1-nuclease protection assays. First an oligonucleotide primer, primer BE05, complementary to a region in the cDNA sequence close to the translation initiation ATG (position −32 to −1; see Figure 4) was used in a primer extension experiment with total RNA purified from adult flies as template. The size of the main product of this reaction was approx. 900 nt; a few shorter products with much less intensity were also apparent on the gel (results not shown). S1-nuclease protection assays using a double-stranded DNA fragment extending from −1 to −1082 (the same primer, BE05, had been utilized to generate the labelled antisense strand; see the Experimental section) gave similar results to the primer extension experiment described above (results not shown). To determine the 5′ end of the mRNA accurately (which might have been detected in the above experiments) another primer, primer BE03, has been utilized (positions −839 to −812). This primer is complementary to a region 818 nt upstream relative to the previous one. This experiment yielded one extension product (Figure 5a). Then S1-nuclease protection assay was performed: the single-stranded radioactively labelled DNA probe complementary to the mRNA was prepared by a 'Dynabeads' solid-phase synthesis method [21] as is described in the Experimental section and is extended from position −818 to −1083. The result of this assay is shown in Figure 5(b). (Both the primer extension and the S1-nuclease protection assay were performed in duplicate.) It is obvious that the two different types of experiments detect the same site as a potential transcription initiation site; the few base differences between the results of primer extension and nuclease protection are not without precedent: e.g. the position of the transcription initiation site of human threonyl-tRNA synthetase determined by these two approaches varies by approx. 10 nt [24]. Slightly shorter extension products by reverse transcriptase may be due to the effect of secondary structure, which cannot be predicted from the primary sequence [25].

The mRNA 5′ end is localized to the adenosyl nucleotide at position −876 (Figure 4). There is a TATA-box (TAATTAT) 23 bp upstream relative to the possible 5′ end of the mRNA (−898 to −892), which can be considered as part of the core promoter.

Comparison of the 5′ end of the *Drosophila* organellar-type Ca^{2+}-ATPase gene to that of the rabbit SERCA2 gene detects a segment with remarkable similarity: the sequence of residues −479 to −466 of the *Drosophila* gene matches at 11 of the 14 positions with the −155 to −142 segment of the rabbit SERCA2 gene (Figure 4).

Analysis of the 5′ region of the *Drosophila* organellar-type Ca^{2+}-ATPase gene revealed potential recognition sequences of various muscle-specific transcription factors (Figure 4). Two identical sequence motifs (AATTCCT), showing similarity to binding site of the M-CAT binding factor (CATTCCT; [26]), can be found at the 5′ end of the gene in positions −1052 to −1047 and positions −1081 to −1076. The M-CAT binding factor (MCBF) recognition sequence first described in the promoter region of the chicken troponin T gene [26] also possesses two copies of this sequence motif separated by 17 bp. The distance of the two motifs in the *Drosophila* Ca^{2+}-ATPase gene is 22 bp. Two MyoD1 binding sites had been identified in the chicken acetylcholine receptor α-subunit gene [27]. The two sites (CCTCAGCTGTC and GGAACAGGTG) are in close proximity, separated by only 6 bp. In the 5′ end of the *Drosophila* Ca^{2+}-ATPase gene two segments can be found similar to the MyoD1 motifs in the chicken acetylcholine receptor α-subunit gene (residues −1039 to −1029 and residues −1020 to −1011). The two sites are 8 bp apart. In addition to the above-mentioned sites the MyoD family consensus sequence (CANNTG) is present in several copies in the 5′ region of the gene. A muscle-specific enhancer recognition sequence, the CArG motif, $CC(A/T)_6GG$,

```
                         ***                                             6760
TTCCTGATGTCGTCGTCGACAGGATGTAGAGGTTTGGACCACCAAAGCCGATACGAATGTGAAACTGAATGTGAAGGAGG
                                                                         6840
ACGAGTTTTCCTGCCTTTATCATTCGTGTATTAATCGCTAAGCCAGCCCTGTTACTTGCAACTGCAACATTAAATCCAAC
                                                                         6920
ATAAATCAATTTATTTGCCTTACAAACTAAAATATTAATGGAACTGTGGGTTAGTAACAACGGATTTaaataaatatata
                                                                         7060
ccaaatatataaattagaacggccctgtttcatttcctttatctgttggaatgggtttatttaataaattagtacagct
                                                                         7120
gaactattctgcggaattacaatatcgacaacgaaattaaagccgccactggaccggaccatttgaccacggagctacga
                                                    7172
gggggggcagcgcgagagggtggccgtcttgtcgcgacgtggagatgaagag
```

Figure 6 Sequence of the 3′ exon and 3′ untranscribed region of the *Drosophila* organellar-type Ca^{2+}-ATPase gene

The exon sequence is capitalized; numbering starts at the translation initiation ATG. The TAG stop codon is shown by asterisks, the polyadenylation signal is double-underlined, and a poly(G/T)-rich sequence 3′ to the polyadenylation site is underlined.

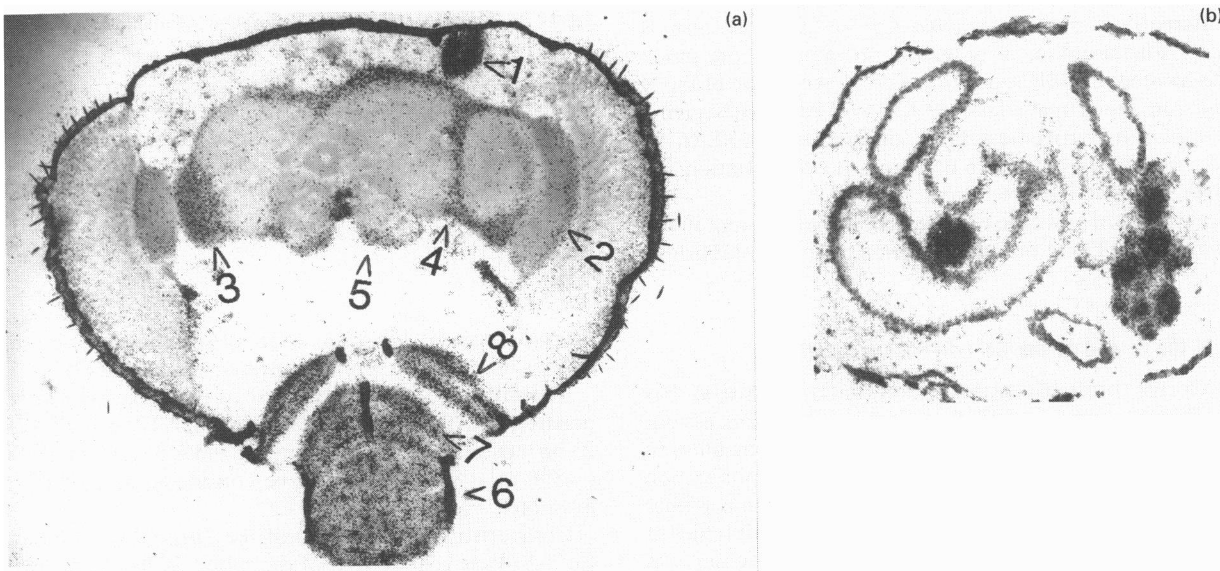

Figure 7 Tissue-specific expression of the organellar-type Ca^{2+}-ATPase gene in *Drosophila melanogaster*

Autoradiographic visualization of *in situ* RNA–RNA hybridization experiments performed on 8 μm micrographs. Antisense ^{35}S-labelled full-size cDNA has been used as hybridization probe at high stringency. (**a**) Vertical slice of an adult head. 1, Ganglion cells of one of the ocelli; 2, those of periopticon; 3, epiopticon; 4, lobus opticus; 5, protocellebellum; 6, muscles of cibarial pump; 7, those of proboscis; 8, laternal pharyngeal muscles. (**b**) Strong hybridization of developing oocytes in the ovarian chamber.

has been described [28]; a very similar sequence is present in the *Drosophila* gene at positions −129 to −120 (GCAAAATAGC). The presence of several possible muscle-specific transcription factor binding motifs in the 5′ end of the gene is in harmony with the results of our *in situ* hybridization experiments which strongly suggest muscle-specific expression of the gene (see below).

The 3′ end of the gene

We have determined the sequence of an additional 214 bp 3′ to the end of the published cDNA sequence (Figure 6). The cDNA-clone sequenced lacked a poly(A) tail [9]. It was found that a polyadenylation signal AATAAA is present in positions 6908 to 6913 with perfect match of the consensus polyadenylation signal sequence [29,30]. The presence of a polyG/T-like sequence 32 bp 3′ to the AATAAA site suggests that a functional polyadenylation signal [30] is located in the 3′ region of the gene.

Tissue-specific expression of the organellar Ca^{2+}-ATPase in *Drosophila*

In order to obtain semi-quantitative data on the tissue distribution of expression of the organellar-type Ca^{2+}-ATPase of *Drosophila melanogaster*, *in situ* RNA–RNA hybridization experiments have been performed on sections (8 μm) of adult flies. The probe used was a ^{35}S-labelled RNA representing the full-size mRNA of the *Drosophila* Ca^{2+}-ATPase [9].

A low mRNA-abundance could be detected in each tissue of adult flies as the grain density in the sections hybridized with the antisense RNA probe was significantly greater than in similar areas of sections hybridized with the sense probe.

The organellar-type Ca^{2+}-ATPase is expressed at a very high level in the central nervous system and in muscles, as is shown in Figure 7(a), which is an *in situ* RNA–RNA hybridization micrograph of an adult fly head. The strong signal of the visual system (in the periopticon, epiopticon opticon and in the optical

lobe), that of the ganglion cells around the ocelli as well as in the antennal centre is apparent on the micrograph. The muscles of the proboscis, the cibarial pump, the laternal pharyngeal muscle and the salivary muscles also possess very high levels of organellar-type Ca^{2+}-ATPase mRNA. The same type of mRNA abundance can be seen in various muscles of the mesothorax and in the tubular leg muscles (results not shown).

As is shown in Figure 7(b), the developing oocytes in the ovarian chamber carry a high amount of organellar-type Ca^{2+}-ATPase mRNA. The egg chamber in *Drosophila* consists of three cell types: the oocyte is connected to its 15 nurse cells and this 16-cell complex is surrounded by follicle cells. The nurse cell nuclei become extremely active in transcription during oocyte development and mRNA of nurse cells is transferred to the oocyte [31]. The oocyte stores about 6000 different mRNA sequences of maternal origin which are utilized in different stages of embryonic development [31]; it was shown that the egg cytoplasm carries a rich supply of maternal products, mRNA and protein, to support early embryogenesis [32]. It is obvious that the organellar-type Ca^{2+}-ATPase mRNA is one of these gene products. The apparently even distribution of the organellar-type Ca^{2+}-ATPase transcript in oocytes may suggest a 'housekeeping' function.

Our data on the tissue-specific distribution of the organellar-type Ca^{2+}-ATPase mRNA are consistent with the hypothesis that this gene serves as a housekeeping one, as a more or less uniform distribution of low-level transcription is observed in the whole adult body. On the other hand high-level transcription in a highly tissue-specific fashion also occurs from the organellar-type Ca^{2+}-ATPase gene. This observed tissue-specificity is in harmony with the physiological role of this type of Ca^{2+} pump (localized in the sarcoplasmic reticulum) in the contraction–relaxation cycle of various muscles such as the rabbit SERCA1a, SERCA1b, SERCA2a and SERCA2b gene products [1,2,4–6] as well as with the central role of Ca^{2+} in signal transduction of nerve cells.

Thanks are due to Dr. A. Udvardy of the Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences for providing the genomic library; to Dr. E. J. Benz, Jr. (Department of Medicine, Yale University School of Medicine, U.S.A.) and to Dr. John Karlson (Department of Biology, Yale University, U.S.A.) for their hospitality during the *in situ* hybridization experiments. We are also thankful for valuable discussions with Dr. J. Szabad (Institute of Biology, A. Szentgyörgyi Medical University, Szeged). The technical help of Györgyi Demeter is also gratefully acknowledged. This work was supported by an OTKA grant (OTKA I/3. 1359) and by a travel fund to A.V. from the Hungarian Soros Foundation.

Received 10 February 1995/18 May 1995; accepted 23 May 1995

REFERENCES

1 MacLennan, D. H., Brandl, C. J., Korczak, B. and Green, N. M. (1985) Nature (London) **316**, 696–700
2 Brandl, C. J., Green, N. M., Korczak, B. and MacLennan, D. H. (1986) Cell **44**, 597–607
3 Burk, S. E., Lytton, J., MacLennan, D. H. and Shull, G. E. (1989) J. Biol. Chem. **264**, 18561–18568
4 Brandl, C. J., de Leon, S., Martin, D. R. and MacLennan, D. H. (1987) J. Biol. Chem. **262**, 3768–3774
5 Lytton, J. and MacLennan, D. H. (1988) J. Biol. Chem. **263**, 15024–15031
6 Guntenski-Hamblin, A.-M., Greeb, J. and Shull, G. E. (1988) J. Biol. Chem. **263**, 15032–15040
7 Campbell, A. M., Kessler, P. B., Sagara, Y., Inesi, G., Shull, G. E. and Fambrough, D. M. (1991) J. Biol. Chem. **266**, 16050–16055
8 Palmero, I. and Sastre, L. (1989) J. Mol. Biol. **210**, 737–748
9 Magyar, A. and Váradi, A. (1990) Biochem. Biophys. Res. Commun. **173**, 872–877
10 Váradi, A., Gilmore-Hebert, M. and Benz, E. J., Jr. (1989) FEBS Lett. **258**, 203–207
11 Escalante, R. and Sastre, L. (1994) J. Biol. Chem. **269**, 13005–13072
12 Escalante, R. and Sastre, L. (1993) J. Biol. Chem. **268**, 14090–14095
13 Korczak, B., Zarain-Herzberg, A., Brandl, C. J., Ingles, C. J., Green, N. M. and MacLennan, D. H. (1988) J. Biol. Chem. **263**, 4813–4819
14 Zarain-Herzberg, A., MacLennan, D. H. and Periasamy, M. (1990) J. Biol. Chem. **265**, 4670–4677
15 Eggermont, J. A., Wuytack, F. and Casteels, R. (1990) Biochem. J. **266**, 901–907
16 Lytton, J. and MacLennan, D. H. (1988) J. Biol. Chem. **263**, 15024–15031
17 Murakami, K., Tanabe, K. and Takade, S. (1990) J. Cell Sci. **97**, 487–495
18 Wimmers, L. E., Ewing, N. N. and Bennett, A. B. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 9205–9209
19 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
20 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
21 Hultman, T., Bergh, S., Moks, T. and Uhlen, M. (1991) BioTechniques **10**, 84–93
22 Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O. and Fields, C. (1992) Nucleic Acids Res. **20**, 4255–4262
23 Bingham, P. M., Chou, T.-B., Mims, I. and Zachar, Z. (1988) Trends Genetics **4**, 134–138
24 Cruzen, M. E. and Arfin, S. M. (1991) J. Biol. Chem. **266**, 9919–9923
25 Boorstein, W. R. and Craig, E. A. (1989) Methods Enzymol. **180**, 347–369
26 Mar, J. H. and Ordahl, C. P. (1990) Mol. Cell. Biol. **10**, 4271–4283
27 Piette, J., Bessereau, J. L., Huchet, M. and Changeux, J. P. (1990) Nature (London) **345**, 353–355
28 Minty, A. and Kedes, L. (1986) Mol. Cell. Biol. **6**, 2125–2136
29 Proudfoot, N. J. and Brownlee, G. G. (1976) Nature (London) **263**, 211–214
30 Wahle, E. and Keller, W. (1992) Annu. Rev. Biochem. **61**, 419–440
31 Spardling, A. C. (1993) in The Development of *Drosophila melanogaster* (Bate, M. and Arias, A. M., eds.), pp. 1–70, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
32 Garcia-Bellido, A. and Moscoso Del Prado, J. (1979) Nature (London) **278**, 346–348