Analysis of the mitochondrial ATP synthase β -subunit gene in *Drosophilidae*: structure, transcriptional regulatory features and developmental pattern of expression in *Drosophila melanogaster*

Pilar PEÑA, Cristina UGALDE, Manuel CALLEJA and Rafael GARESSE*

Departamento de Bioquímica (UAM), Instituto de Investigaciones Biomédicas (CSIC), Facultad de Medicina, Universidad Autónoma de Madrid, c/ Arzobispo Morcillo 4. 28029 Madrid, Spain

We have cloned and determined the structure of the gene encoding the H^+ -ATP synthase β subunit in two distantly related *Drosophila* species, *D. melanogaster* and *D. virilis*. The gene contains three exons that are extremely well conserved at the amino acid level, not only in the region encoding the mature protein but also in that encoding the leader peptide. Primer extension analysis indicates that the 5' untranslated region is extremely short, and reveals the presence of multiple initiation sites of transcription in both *Drosophila* species. The promoters of *D. melanogaster* and *D. virilis* H^+ -ATP synthase β -subunit genes contain a conserved region surrounding the initiation

transcription sites. Nucleotide sequence analysis has revealed the absence of canonical TATA and CCAAT boxes and the presence of several putative regulatory elements in both promoter regions, including GAGA, GATA and Ets binding sites. We have analysed the pattern of gene expression during *D. melanogaster* development. The mRNA is stored in oocytes, and activation of transcription takes place after 10 h of development. The expression of the nuclear-encoded H^+ -ATP synthase β subunit is strictly coordinated with the expression of subunits 6 and 8 of the same complex that are encoded in the mitochondrial genome.

INTRODUCTION

Mitochondrial ATP synthase (F_1F_0 -ATPase) catalyses the synthesis of ATP using the electrochemical gradient across the inner mitochondrial membrane generated during oxidative phosphorylation. It is composed of two functional and structurally well-defined parts, the proton channel F_0 embedded in the membrane, and the hydrophilic F_1 , which contains the catalytic activity. The entire F_1F_0 complex contains 12–14 subunits, two of them (ATPase 6 and 8) encoded in the mitochondrial genome. The rest, encoded in the nucleus, are synthesized in the cytosol and imported into the mitochondria [1–3]. F_1 consists of five subunits with the stoichiometry $\alpha_3\beta_3\gamma\delta\epsilon$. Recently, its tri-dimensional structure has been elucidated at 2.8 Å resolution [4]. The two major subunits of the F_1 complex, α and β , contain the nucleotide-binding sites and play an essential role in the catalytic function of the enzyme.

Different physiological conditions, such as metabolic changes, cell proliferation, differentiation and development, require the active synthesis of mitochondrial proteins [5–7]. On the other hand, different organs and tissues have different respiratory demands and therefore need different mitochondrial functions [8]. Since the biogenesis of mitochondria depends on products encoded in two different genomes, to adapt mitochondrial function to the variable physiological needs of the organism a cross-talk between nuclear and mitochondrial DNA is required. However, our current understanding of the mechanisms controlling the coordinated expression of mitochondrial genes is very limited. The recent characterization of several mammalian nuclear-encoded mitochondrial genes has allowed the identification in its 5' proximal promoter regions of some motifs that could play a regulatory role in mitochondrial biogenesis and two

transcription factors, designated Nuclear Respiratory Factors 1 and 2 (NRF-1, -2) have been purified from HeLa cells [9,10]. It has been recently suggested that NRF-1, initially characterized as an activator of the rat cytochrome *c* gene [11], plays a broad role in nucleo-mitochondrial interaction [9,12]. Among the genes regulated by NRF1 are those encoding the regulatory transcription factor mtTFA, the RNA component of the mitochondrial RNA processing enzyme (MRP) and some genes encoding proteins important for mitochondrial biogenesis, for example the 5-aminolevulinate synthase, the key enzyme in the biosynthesis of the haem group [9]. NRF-1 has been cloned and its deduced amino acid sequence shows extensive homology in the C-terminal region with the products of two genes that are important in the development of *Drosophila* and sea urchin, ERECT WING and P3A2 respectively [9]. NRF-2, which also regulates the expression of several mitochondrial genes, has also been cloned and is homologous to mouse GABP (GA-binding protein), a member of the Ets family of transcription factors [10,13]. Ets proteins regulate gene expression during a variety of biological processes and have been characterized in several organisms, ranging from fly to man. They interact with the DNA through the Ets domain, first identified in the viral *ets* oncogene [14]. Much less is known in non-mammalian organisms. Among them, *Drosophila* offers an excellent opportunity to combine genetic and molecular techniques, a strategy that has been successfully applied in the study of yeast mitochondrial biogenesis [2,15]. As a first step to studying the regulation of the expression of nuclear-encoded mitochondrial genes in *Drosophila* we have cloned the gene encoding the β subunit of the H^+ -ATP synthase complex (β -ATPase). Although many cDNAs from different organisms encoding the β -ATPase have been analysed, the structure of the gene has been determined in only a few cases.

Abbreviations used: β -ATPase, subunit β of mitochondrial H^+ -ATP synthase; ATPase 6-8, subunits 6 and 8 of mitochondrial H^+ -ATP synthase; NRF-1 and -2, Nuclear Respiratory Factors 1 and 2.

* To whom correspondence should be addressed.

The sequences presented in this paper have been deposited in the EMBL database under the accession numbers X86015, X86016 and X86017.

Figure 1 Nucleotide sequence of the gene encoding the *D. melanogaster* β -ATPase subunit and predicted amino acid sequence of the exons

The nucleotide sequence is numbered from the first base (A) of the codon specifying the initiator methionine (+1). The exon/intron borders are indicated by arrows. Amino acids are shown using the one-letter code, with the termination code denoted by an asterisk. Putative binding sites for transcription factors in the 5' upstream region are underlined with the name of the motifs shown above the sequence when encoded in the sense strand or below the sequence when encoded in the antisense strand. The sequence has been entered in the EMBL database under the accession number X86015.

Specifically in animals only the human β -ATPase gene has been characterized until now [16,17]. In this paper we describe the cloning and characterization of the β -ATPase gene in two distantly related *Drosophila* species, *D. melanogaster* and *D. virilis*. The gene is organized in three exons interrupted by two introns and its expression is strictly coordinated with that of the mitochondria-encoded ATPase 6 and 8 genes during development. The comparison of the 5' regions of the *D. melanogaster* and *D. virilis* β -ATPase genes has revealed the presence of some conserved and non-conserved features that could be important in understanding how the expression of the gene is controlled.

MATERIALS AND METHODS

Library screening and isolation of clones

Drosophila melanogaster genomic λ -EMBL-3 and λ -DASH libraries (kindly provided by Dr. K. Kaiser and Dr. C. Gonzalez respectively) were screened with a *Drosophila* β -ATPase cDNA probe [18]. Then 3×10^5 plaques from each library were transferred to nitrocellulose filters, prehybridized in 40% formamide/6 $\times$ SSC/1% SDS/5 $\times$ Denhardt's solution/100 μ g/ml of denatured salmon sperm DNA at 42 $^{\circ}$ C (SSC: 0.15 M NaCl/0.015 M sodium citrate) and then hybridized overnight in the same conditions with the *Drosophila* probe labelled with [α - 32 P]dCTP. Filters were washed at high stringency (0.5% SDS/0.1 $\times$ SSC at 68 $^{\circ}$ C) and autoradiographed with intensifying screens at -70 $^{\circ}$ C. Positive clones were purified by two additional rounds of screening, and the phages amplified using standard protocols [19].

To clone the *D. virilis* β -ATPase gene, 3×10^5 plaques from a λ -DASH genomic library from this organism (kindly provided by Dr. S. Campuzano) were screened and the phages purified in the same conditions described above.

DNA sequencing

The nucleotide sequence was determined using the dideoxy chain-termination method [20] with Sequenase (USB) and polyacrylamide gels or *Taq* polymerase and automatic sequencing (373 DNA sequencer, Applied Biosystems) following the manufacturers' instructions. The DNA sequence was obtained mostly in both directions using a shotgun strategy [21] with every nucleotide position determined four times on average. Sequences were analysed using the programs developed by Staden [22] and the GCG programs of the University of Wisconsin [23] on a Digital Vax computer.

Hybridization to polytene chromosomes

Third-instar larval salivary glands, squashed as described in [24], were hybridized with cDNA or genomic *Drosophila* β -ATPase probes labelled with biotin in 8 $\times$ SSC containing 2 $\times$ Denhardt's solution, 20% dextran sulphate and 0.4% salmon sperm DNA for 10 h at 58 $^{\circ}$ C. The slides were washed three times in 2 $\times$ SSC (20 min each), twice in PBS (5 min each) and once in PBS/0.1% Triton X-100 (2 min) and finally rinsed in PBS. After washing, the squashes were incubated for 30 min at 37 $^{\circ}$ C with fluorescein avidine D (Vector Laboratories) diluted 1:50 in PBS, developed with 10 mg/ml diaminobenzidine, and stained with 0.5 μ g/ml of Propidium Iodide in PBS. The squashes were mounted in 85% glycerol containing 2.5% of n-propyl gallate, and analysed and photographed using confocal microscopy.

Determination of the initiation of transcription

For primer extension analysis, 5 pmol of different specific oligonucleotides were labelled with 50 μ Ci of [γ - 32 P]ATP and polynucleotide kinase. Aliquots (2 μ g) of poly(A) or 25–40 μ g of total RNA and 200 000 c.p.m. of 32 P-labelled primer were used in each experiment. Hybridizations were carried out at 65 $^{\circ}$ C and the primer extended with 50 units of avian myeloblastosis virus (AMV) reverse transcriptase for 90 min at 37 $^{\circ}$ C. The extended products were analysed on 8% polyacrylamide/7 M urea gels. Sequencing reactions using the same oligonucleotide were run in parallel.

For RNase protection assays, a 0.9 kb *Clal*–*EcoRI* genomic fragment (from position -476 to +425 in Figure 1) was subcloned in the plasmid Bluescript (Stratagene) and transcribed *in vitro* with T7 RNA polymerase in the presence of [32 P]GTP in order to obtain the antisense RNA. The RNase protection analysis was carried out using the RPA2 kit from Ambion following the manufacturer's indications.

RNA extraction and Northern blot analysis

Oocytes and timed collections of embryos, larvae and adults were obtained from Oregon-R wild-type *D. melanogaster* stocks. Total RNAs were extracted following a protocol described previously [19]. For Northern blot hybridizations, total RNA was electrophoresed on 1.2% agarose/1.8 M formaldehyde gels, blotted on to a nylon membrane and probed with the β -ATPase cDNA or an ATPase 6-8-specific clone labelled with [α - 32 P]dCTP. The filters were probed, washed and autoradiographed in the same conditions as described above for the library screenings.

RESULTS

Structure and chromosome mapping of the *D. melanogaster* β -ATPase gene

We previously reported the cloning and sequencing of a cDNA encoding the *D. melanogaster* ATPase β subunit [18]. To isolate the corresponding gene, we used this cDNA as probe to screen genomic libraries as described in the Materials and methods section, and several independent positive clones were identified and plaque-purified. Standard restriction enzyme analysis using the enzymes *SalI*, *BamHI* and *EcoRI* showed that they corresponded to the same genomic region. One of them (λ -DMB4) was selected for a detailed characterization, and the DNA fragments that hybridize with the β -ATPase cDNA probe were identified by Southern blot analysis. These fragments, covering a region of roughly 5 kb, were subcloned in Bluescript and sequenced, showing that they contain the complete gene as well as 5'- and 3'-flanking sequences (Figure 1). The number and position of introns within the gene were identified by direct comparison of the genomic and cDNA sequences. The gene contains two introns; the first, located in the middle of the region encoding the presequence, is 956 bp long. The second is located inside the region encoding the mature protein and is 52 bp long (Figure 1). In both cases the sequences of the intron/exon boundaries are strictly conserved and correspond to the consensus described for eukaryotic introns: GTAAAGT (donor site)/CAG (acceptor site). The sequences of the exons were 100% identical to those of the corresponding regions of the *D. melanogaster* β -ATPase cDNA.

Using the β -ATPase cDNA, as well as genomic clones, as probes we have localized the gene in the polytene chromosomes of *D. melanogaster* as described in the Materials and methods section. There is only a single hybridization signal which maps

Figure 2 *In situ* hybridization of a β -ATPase probe to salivary gland polytene chromosomes

A confocal photograph of the squashed polytene chromosomes is shown in (a). (b) shows the hybridization of these chromosomes to the β -ATPase probe. A single hybridization signal is detected in the 102D region of the fourth chromosome (indicated by an arrow).

in the band 102D of the fourth chromosome (Figure 2). This result is in accordance with Southern blot analysis (results not shown), showing that in *D. melanogaster* there is only a single gene encoding the β -ATPase subunit.

Transcription initiation sites and structure of the 5' upstream region

The site of initiation of transcription was determined by primer extension analysis using total RNA extracted from *D. melanogaster* embryos and two different oligonucleotides Dm1 and Dm2 (indicated in Figure 3a). The transcription initiation sites calculated using both oligonucleotides were identical and the result obtained using Dm2 is shown in Figure 3(b). Several extended fragments are clearly visible at positions 12, 16, 25 (first start region) and 47 and 50 (second start region) from the initiator methionine. The strongest signals correspond to the positions located at 12, 16 and 25.

The initiation of transcription was confirmed by RNase protection assays as described in the Materials and methods section. The results are shown in Figure 4(a), where the DNA fragment used as probe is also indicated. Two main protected fragments are detected, corresponding to start points located at positions 12 and 16. Fainter bands are also visible at longer exposition of the gels in equivalent positions to those detected with primer extension analysis (results not shown). A summary of the results obtained with the two techniques is schematically shown in Figure 4(b), where the positions corresponding to the faint bands detected by RNase protection have also been included.

Canonical TATA and CCAAT boxes were not found at 25 to 30 bp and 50 to 100 bp respectively from these initiation sites. There are, however, two short AT-rich regions, TAAATA and AATATAT, located at positions 85 to 80 and 76 to 70 respectively from the initiation of translation that could be used as alternative TATA boxes in the second start region. The sequence analysis of the two initiation regions reveals the presence of the pentanucleotide TCAGT (first start region) and the related sequence TCACT (second start region; see Figure 4b). It has been

Figure 3 Primer extension analysis of the *D. melanogaster* β -ATPase transcription initiation sites

(a) Proximal 5' upstream nucleotide and predicted amino acid sequences of the *D. melanogaster* β -ATPase gene. The positions of the two oligonucleotides (Dm1 and Dm2) used in the analysis are shown. The position of the first intron of the gene is indicated by an arrow. (b) Samples (40 μ g) of total RNA from overnight embryos were hybridized to 32 P-labelled oligonucleotide Dm2 and incubated with avian myeloblastosis virus reverse transcriptase. The extended products were analysed in a 8% polyacrylamide/7 M urea gel. The sequence obtained using Dm2 as primer is presented, and the positions of the main initiation sites are indicated on the left of the Figure.

suggested that the eukaryotic CAP-site motif TCAGT [25] plays an important role in transcription [26].

In order to identify potential regulatory elements we have sequenced 2 kb of the 5' upstream region and searched for the presence of putative eukaryotic protein-binding sites. In Figure 1, the positions of several sequence motifs recognized by transcription factors are indicated. They include CCAAT, SP1, and AP3 binding sites, and several motifs matching the consensus recognized by the GATA and GAGA regulatory proteins. In addition there are several CANNTG (E-boxes) recognized by the muscle-specific transcription factors of the myoD family [27] and a putative binding site for the myogenic factor MEF-2, that has been recently characterized in *Drosophila* [28].

Finally, to explore the possibility that the 5' region of the *D. melanogaster* β -ATPase gene contains elements that could be

Figure 4 RNase protection analysis of the *D. melanogaster* β -ATPase transcription initiation sites

(a) The upper panel indicates the position of the probe used in the RNase protection analysis. It extends from the *Clal* (position -476) to the *EcoRI* (position +425) sites. The regions encoding the presequence are represented by stippled boxes, and the region encoding the 5' end of the mature protein by an open box. The position of the oligonucleotide Dm2, complementary to the 3' end of the first exon (see Figure 3a) is shown by an arrow. In the lower panel the result obtained in the RNase protection assay is shown. In lane (+) 50 μ g of total RNA obtained from overnight embryos was hybridized to the 32 P-labelled single-stranded probe, digested with RNase as described in the Materials and methods section, and the protected fragments analysed in a 8% polyacrylamide/7 M urea gel. In lane (-) a control with yeast RNA was included. The sequence has been obtained using the oligonucleotide Dm2 as primer. The protected fragments correspond to the start sites located at -12 and -16. (b) The transcription start sites determined by primer extension and RNase protection analysis are indicated by arrows. The positions corresponding to faint bands detected after long exposure of the RNase protection assay gels (25, 47 and 50) have also been included. The CAP-site motif TCAGT and the related sequence TCACT are underlined.

recognized by other factors similar to that controlling the expression of vertebrate mitochondrial genes, we have also searched for the presence of sequences matching consensus binding sites for NRF-2 [10], NRF-1 [9], REBOX/OXBOX [29] and Mt [30] factors. The transcription factor NRF-2 is homologous to the mouse GABP protein [10], a member of the Ets family, that recognizes the trinucleotide motif GGA [14]. On the other hand, the expression of the human β -ATPase gene has been recently shown to be controlled by Ets proteins that interact with a tandem of Ets-binding sites (core consensus A/CGGAAG/C) located in the proximal promoter region of the gene [31]. Interestingly, in the 5' region of the *D. melanogaster* β -ATPase gene there are eight motifs matching this core consensus binding site, some of them located in tandem (positions from 544 to 557 and from 1298 to 1312 in Figure 1). NRF-1 (or α -Pal, see [32])

has a DNA-binding domain conserved in the sea urchin P3A2 and *Drosophila* ERECT WING proteins [9]. The binding site for NRF-1 has been characterized and corresponds to the consensus TGCGCATGCGCA, where the tetranucleotide GCGC is the site of contact with the transcription factor. The binding site for P3A2 has also been determined [33] and corresponds to T/CNT/CGCGCA/T, approximately half of the NRF-1 binding site. In the 5' region of the *D. melanogaster* β -ATPase gene there are three boxes with a 5 to 6 (positions 1177 to 1172 and 1553 to 1548) or 6 to 6 (position 1293 to 1288) identity to the core consensus T/CGCGCA/T (Figure 1) and which are therefore candidates to be recognized by factors of the NRF-1 family. There are no elements with significant identity to the OXPHOX/REBOX complex or the Mt-binding sites.

Gene structure of the *D. virilis* β -ATPase gene

It is widely admitted that during the course of evolution only functionally important elements should exhibit a significant conservation. With the aim of delimiting putative regulatory elements in the 5' control region of the β -ATPase gene, we have used an evolutionary strategy that has been successfully applied to study the promoter region of several *Drosophila* genes [34–37], i.e. to compare the gene structure and organization in different *Drosophila* species. In this context we have cloned and analysed the gene encoding the β -ATPase in *D. virilis*, a species that diverged from *D. melanogaster* 60 million years ago [38]. A *D. virilis* genomic library prepared in λ -DASH was screened using cDNA and genomic *D. melanogaster* β -ATPase clones as probes and several positive phages were identified and purified as described in the Materials and methods section. One of them was analysed in detail and the DNA fragments containing the β -ATPase gene were identified by Southern blot analysis, subcloned in Bluescript, and after partial sequencing of these subclones the exons encoding the gene were identified. The structure of the gene is strictly conserved with respect to its *D. melanogaster* counterpart and is shown schematically in Figure 5(a). The exons have a similar size in both *Drosophila* species; the first intron is roughly 400 bp shorter, and the small second intron is 20 bp longer in *D. virilis*. The partial sequences obtained corresponding to the coding region (Figure 5b) show a high degree of conservation between the two *Drosophila* species, which is apparent in both the mature protein and the signal peptide. There is a 95% identity at the amino acid level and a lower identity (81%) at the nucleotide level, mainly due to conservative changes in third codon positions (Figure 5b).

To localize the transcription start site, primer extension analysis was carried out using total RNA extracted from *D. virilis* imagos and two different oligonucleotides, Dv1 and Dv2, located in the positions indicated in Figure 6(a). The transcription initiation sites calculated using both oligonucleotides are coincident, and the result obtained using Dv2 is shown in Figure 6(b). Several extended fragments are present, indicating that the *D. virilis* gene has an heterogeneous initiation of transcription also. The strongest signal is also located very close to the initiator methionine (position -13) and fainter bands are located at positions -16, -26, -34 and -35. The analysis of the sequence revealed that the CAP-site motif TCAGT is present at position -14 to -10. Similar to findings in the *D. melanogaster* β -ATPase gene, there are not TATA or CCAAT boxes located in canonical positions 5' upstream from these initiation sites, although there are A/T-rich sequences located at positions -52 to -48 (ATATT) and -66 to -62 (TAAAT) that could be used as alternative TATA boxes.

Figure 5 Structure of the *D. melanogaster* and *D. virilis* β -ATPase genes, and sequence comparison of part of the protein-coding regions

(a) The gene structure of the *D. melanogaster* and *D. virilis* β -ATPase genes is schematically represented. Filled boxes represent the regions encoding the presequence, stippled boxes the regions encoding the mature protein and open boxes the introns. The regions of the *D. virilis* gene for which the nucleotide sequence have been determined are indicated below the scheme with arrow-ended solid lines. (b) Sequence comparison of the *D. melanogaster* and *D. virilis* protein-coding regions corresponding to the first, second and part of the third exons of the β -ATPase gene. The nucleotide and predicted amino acid sequences presented correspond to the *D. virilis* gene. Amino acids are represented by the one-letter code. The nucleotides and amino acids differing in the *D. melanogaster* gene are shown respectively below and above the *D. virilis* sequence. Splice sites are indicated with thick arrows and the leader peptide processing site by a thin arrow. The sequence has been entered in the EMBL database under the accession number X86017.

We have sequenced the 5' upstream proximal region from the initiation of translation (roughly 800 bp, Figure 7a) of the *D. virilis* β -ATPase gene. Sequence comparison with the *D. melanogaster* counterpart region has revealed that the overall homology between these two equivalent regions is very low, making it impossible to align them. The only exception is a short region located adjacent to the initiator methionine encompassing the

Figure 6 Primer extension analysis of the *D. virilis* β -ATPase transcription initiation sites

(a) Proximal 5' upstream nucleotide and predicted amino acid sequences of the *D. virilis* β -ATPase gene. The positions of the two oligonucleotides used in the analysis are shown. (b) 25 μ g of total RNA extracted from adults was hybridized to 32 P-labelled oligonucleotide Dv2 and incubated with avian myeloblastosis virus reverse transcriptase. The extended products were analysed in 8% polyacrylamide/7 M urea gels. The nucleotide sequence obtained using the Dv2 as primer is presented, and the positions of the initiation sites are indicated on the right of the Figure.

short non-translated region and 5' proximal upstream sequences. It is 73 nucleotides long in *D. melanogaster* and 52 nucleotides long in *D. virilis* due to the presence of two deletions of 13 and 10 nucleotides respectively and a single insertion of two nucleotides (Figure 7b). Excluding these insertion/deletions this element shows an 80% identity. The search for putative protein-binding sites in the 5' region of the *D. virilis* β -ATPase gene has revealed the presence of several motifs that are also present in the equivalent region of the *D. melanogaster* gene. There are several Ets, GAGA, GATA and CANNTG boxes that are underlined in Figure 7(a). In the position -727 to -722 there is also an element conserved to the P3A2 (NRF1-like)-binding site (Figure 7a). Therefore, although the overall homology is very low, both 5' regions show a similar organization with respect to the presence of putative elements recognized by regulatory proteins.

Developmental pattern of expression

We have determined by Northern blot analysis the pattern of expression of the nuclear-encoded β -ATPase and the mitochondrial-encoded ATPase 6 and 8 genes during the development of *D. melanogaster*. Total RNA extracted from oocytes, embryos of different developmental stages, first-stage

Figure 7 5' control region of the *D. virilis* β -ATPase gene and comparison with its *D. melanogaster* counterpart

(a) The 800-nucleotide-long sequence corresponding to the 5' region of the *D. virilis* β -ATPase gene is presented. Position +1 corresponds to the first base (A) of the codon specifying the initiator methionine. Putative binding sites for transcription factors are underlined with the names of the motifs shown above the sequence when encoded in the sense strand or below the sequence when encoded in the antisense strand. (b) Comparison of the *D. melanogaster* (Dm) and *D. virilis* (Dv) 5' proximal regions. Dashes represent gaps introduced to maximize identity. Equivalent nucleotides are denoted by asterisks (*). The position +1 corresponding to the first nucleotide of the coding region is indicated. The sequence has been entered in the EMBL database under the accession number X86016.

larvae and imagoes were electrophoresed on agarose/formaldehyde gels, transferred to nylon membrane and probed with a 32 P-labelled β -ATPase probe as described in the Materials and methods section. After washing, the filter was reprobed with a specific ATPase 6-8 probe and a cytoplasmic ribosomal probe to check the loading of the gel. The results are presented in

Figure 8, showing that both messenger RNAs (β -ATPase and ATPase 6-8) are stored in oocytes. Their steady-state levels decrease during the first hours of development, and increase later on (between 6 and 12 h), reaching a plateau that is maintained in larvae and imagoes. Although the decrease is somehow slower for the ATPase 6-8 transcript, the pattern of gene expression is

Figure 8 Northern blot analysis of the steady-state level of β -ATPase and ATPase 6-8 transcripts during *D. melanogaster* development

Samples (10 μ g) of total RNA from oocytes (0o), embryos of the indicated time, first-stage larvae (L1) and adults (A) were electrophoresed on a 1.2% agarose/1.8 M formaldehyde gel, transferred to a nylon membrane and sequentially hybridized with the 32 P-labelled cDNA β -ATPase clone and a specific probe for the ATPase 6-8 genes (an *Xba*I *D. melanogaster* mtDNA fragment from positions 3349 to 4316 as numbered in [58], subcloned in Bluescript). To check the loading of the gel the filter was rehybridized with an 18 S cytoplasmic rRNA probe (Pdm R).

similar for both nuclear- and mitochondria-encoded genes, that therefore show a coordinated response.

DISCUSSION

We have determined the structure and the complete sequence of the gene encoding the H^+ -ATP synthase β -subunit in *D. melanogaster* as well as its 5' and 3' proximal regions. From another evolutionarily relatively distant *Drosophila* species, *D. virilis*, the structure and partial nucleotide sequences of the H^+ -ATP synthase gene have also been obtained. This characterization reveals that it is organized in three exons separated by two introns, covering a genomic region of roughly 3 kb. The gene organization is strictly conserved in the two *Drosophila* species; the first intron (about 1 kb long in *D. melanogaster* and 0.6 kb long in *D. virilis*) is located in the middle of the presequence, and the second, located in the region encoding the mature protein, is 72 bp long in *D. virilis* and 52 bp long in *D. melanogaster*, one of the smallest intron sizes described to date [39]. The low number of introns in the *Drosophila* β -ATPase gene contrasts with the high number found in the β -ATPase genes characterized in other organisms: seven in *Neurospora crassa* [40], nine in *Nicotiana tabacum* [41] and 10 in human [16]. The only exception is *Saccharomyces cerevisiae* [42], whose β -ATPase gene has no introns, like most of the genes characterized in this organism. The positions of the *Drosophila* introns are not conserved in other β -ATPase genes.

In accordance with its central role in ATP synthesis, the amino acid sequence of the β -ATPase subunit has been highly conserved during evolution. For example the *Drosophila* mature protein has a 92% identity with the human, and 81% identity with the yeast β -ATPase subunits [18]. The comparison between the sequenced regions corresponding to the mature protein in *D. virilis* with the equivalent regions in *D. melanogaster* reveals a high conservation at the amino acid level (95%). Interestingly, the conservation is also very high in the presequence, where only one amino acid change (serine/alanine) has been detected. There is a large amount of evidence showing that the leader sequence of mitochondrial proteins is highly variable, both in length and primary sequence, and the conservation is restricted to its basic character and secondary structure [43]. For example, the β -ATPase presequences from human, yeast and *Drosophila* are of variable size and almost completely different in sequence [18].

Although the reason for the high amino acid conservation detected between the β -ATPase presequence of the two *Drosophila* species is not clear, it may be related to the conservation of specific components of the mitochondrial import machinery with which this protein has to interact.

Using β -ATPase clones as probes, we have detected a single hybridization signal mapping in the position 102D of the fourth polytene chromosome of *D. melanogaster*. Several additional arguments also indicate that the β -ATPase subunit is encoded by a single copy gene, a situation also found in all the systems studied to date: all the phages isolated in exhaustive library screenings correspond to the same genomic region, in Southern blot experiments only the bands corresponding to this region are detected (results not shown) and Northern blot experiments reveal the presence of one single RNA of 1.7 kb [18].

We have determined the initiation site for transcription of the *D. melanogaster* β -ATPase gene using two different techniques, primer extension and protection from RNase digestion. The results obtained are coincident and reveal a heterogeneous initiation of transcription in two main regions, separated by 30 nucleotides, each of them with several start points. The absence of TATA and CAATT boxes in canonical positions indicate that the gene has a TATA-less promoter. Both features, multiple transcription start points and absence of a TATA box, are usually associated with housekeeping genes. In particular, the human β -ATPase gene also contains a TATA-less promoter and, depending on the authors, has one or two points of transcription initiation [16,17]. One conserved feature in both *D. melanogaster* start regions is the presence of the pentanucleotide TCAGT, which is also present in the main start region mapped in the homologous human gene. The sequence TCAGT has been named the 'arthropod initiation site' and probably plays an important role in transcription [26]. In *D. virilis*, we have found a pattern very similar to that described above for the *D. melanogaster* homologous gene, with a very short untranslated region, a heterogeneous initiation of transcription, lack of canonical TATA and CAATT boxes and presence of the CAP-site motif TCAGT. These characteristics seem therefore to be conserved in the β -ATPase gene of Drosophilidae.

The structural characteristics associated with the non-translated regions of eukaryotic cytoplasmic mRNAs have been studied by Kozac [44,45]. In addition to the nucleotide sequence surrounding the AUG initiator codon, the lack of secondary structure of the region (low G+C content) and a moderated long leader sequence are important for translational efficiency [45]. The leader length of the *Drosophila* β -ATPase gene is extremely short, close to the minimum of 10 nucleotides permissible, ranging between 12 and 50 nucleotides in *D. melanogaster* and between 13 and 35 nucleotides in *D. virilis* depending on whether we consider the more proximal or the more distal start points. The influence of this feature in the translational efficiency of the β -ATPase gene is unknown.

We have sequenced the 5' proximal region of both β -ATPase genes, which probably encodes elements important for the control of gene expression. A similar strategy has been used before to identify functionally relevant conserved regions in the promoter of several *Drosophila* genes. Comparison of the 5' regions of *D. melanogaster* and *D. virilis* *twist* [34] and *engrailed* [35] genes or *D. melanogaster* and *D. hydei* *fushi tarazu* [36] genes has revealed a considerable conservation in several relatively long regions to which has been ascribed an important role in transcriptional regulation. The 5' control region of the *Ultrabithorax* gene also has several regions well conserved among *D. pseudoobscura*, *D. funebris*, *D. melanogaster* and *Musca domestica* [37]. Surprisingly, the 5' region of the *D. melanogaster* and *D. virilis* β -ATPase

genes has diverged considerably. The only element relatively well conserved is the region located adjacent to the initiation of translation, containing the transcription start points and surrounding sequences, suggesting that this region should play a crucial role in the transcriptional regulation of this gene.

We have analysed the proximal 5' region of the *D. melanogaster* and *D. virilis* β -ATPase genes, searching for the presence of putative regulatory elements. We have detected the presence of Ets-binding sites, a sequence element which has particular relevance because of their functional role as transcriptional activators in some mammalian nuclear-encoded mitochondrial genes, including the human β -ATPase gene [31]. We have also detected several elements showing a considerable identity to the NRF-1 binding site. The NRF-1 DNA-interacting domain is conserved in the protein encoded by the *Drosophila* gene *erect wing* [9,46]. Genetic analysis of the *erect wing* locus has shown that it plays an essential role in muscle and nervous system differentiation [47]. Interestingly, the most evident phenotype of some *erect wing* mutants is hypoactivity [48], an abnormality that could be related to a decrease in mitochondrial energy production. The finding in the 5' upstream region of the *Drosophila* β -ATPase gene of putative Ets- and NRF-1-binding sites opens the interesting possibility that similar factors participate in the control of nuclear-encoded mitochondrial genes in vertebrates and invertebrates.

The comparative sequence analysis has also revealed the presence in both 5' regions of several putative motifs recognized by GATA and GAGA proteins, two transcription factors that play a broad role in transcriptional activation. The GATA proteins consists of a family of transcription factors that recognize the binding site WGATAR and have been found in a variety of organisms, including *Drosophila* [49,50]. The GAGA proteins recognize the core tetranucleotide sequence GAGA, a binding site that is present in the promoter regions of several *Drosophila* genes. There are several examples showing the importance of the GAGA factor for achieving maximal levels of gene expression, probably preparing the promoter for activation mediated by other transcription factors [51,52]. Although the presence of similar putative protein-binding motifs in the 5' region of the *D. melanogaster* and *D. virilis* β -ATPase genes is quite suggestive, especially in light of the low level of overall conservation of this region, functional studies are needed to establish the possible physiological role of these elements.

We have finally studied the expression of the β -ATPase gene during the development of *D. melanogaster*. During development there are substantial changes in mitochondrial maturation and mitochondrial enzyme activities. Studies carried out in mouse [53], sea urchin [54], *Xenopus* [55] and fish [56] suggest that the expression of the mitochondrial genome is controlled at the level of transcription, affecting differentially ribosomal and messenger RNAs. We have evaluated, during the developmental changes in the steady state, levels of two mitochondrial transcripts that participate in the biogenesis of the ATP synthase complex, the nuclear-encoded β -ATPase and the mitochondrial-encoded ATPase 6-8. To our knowledge this is the first time that the levels of nuclear and mitochondrial genes encoding proteins of the same complex have been quantified during development of any organism. In *Drosophila* the ATPase 6 and 8 genes overlap, and probably are transcribed in a single transcript as has been reported for other systems.

Both β -ATPase and ATPase 6-8 transcripts are stored in *Drosophila* oocytes. During the first stages of development there is a continuous decrease in the steady-state level of both transcripts, more pronounced in the case of β -ATPase (probably due to a lower stability of the nuclear-encoded mRNA compared

with the ATPase 6-8 transcript). At around 6 h of development the steady-state levels of both transcripts are very low, and their amounts increase coordinately as development proceeds, reaching a plateau that is maintained in first-stage larvae and adults.

Drosophila mitochondrial DNA replication has been previously studied [57]. Mitochondrial DNA is present in oocytes in high amounts representing 50 % of total DNA, and it does not start replication until after 10 h of development [57]. Our results indicate that the activation of mitochondrial gene expression occurs between 6 and 12 h of development, suggesting that may be coupled to the process of mitochondrial DNA replication. In addition, the activation of transcription of nuclear-encoded and mitochondrial-encoded genes is coordinated. This coordination would probably require the presence of some factors interacting with both the mitochondrial and the nuclear genomes. The identification and characterization of these putative factors will be one of the main objectives for the near future in the field of mitochondrial research.

We wish to thank Drs. K. Kaiser, C. Gonzalez and S. Campuzano for the generous gifts of *Drosophila* libraries, and Dr. C. Gonzalez for his invaluable help in the mapping of the gene to *Drosophila* polytene chromosomes. We also thank Leandro Sastre and Roberto Marco for critical reading of the manuscript. The expert technical assistance of Pilar Ochoa is fully acknowledged. P.P. was the recipient of a predoctoral fellowship from the Fondo de Investigaciones Sanitarias de la Seguridad Social (FISs) and C.U. was a recipient of a fellowship from Dornier (Germany). This work has been supported by grants of the Comunidad Autónoma de Madrid 131/92 and FISs 93/0339.

REFERENCES

- Senior, A. E. (1988) *Physiol. Rev.* **68**, 177–231
- Attardi, G. and Schatz, G. (1988) *Annu. Rev. Cell Biol.* **4**, 289–333
- Nagley, P. (1991) *Trends Genet.* **7**, 1–3
- Abrahams, J. P., Leslie, A. G. V., Lutter, R. and Walker, J. (1994) *Nature (London)* **370**, 621–628
- Williams, R. S. (1986) *J. Biol. Chem.* **261**, 12390–12394
- Valcarce, C., Navarrete, R. M., Encabo, P., Loeches, E., Satrustegui, J. and Cuezva, J. M. (1988) *J. Biol. Chem.* **263**, 7767–7775
- Soboll, S. (1993) *Biochim. Biophys. Acta* **1144**, 1–16
- Wallace, D. C. (1992) *Annu. Rev. Biochem.* **61**, 1175–1212
- Virbasius, C. A., Virbasius, J. V. and Scarpulla, R. C. (1993) *Genes Dev.* **7**, 2431–2445
- Virbasius, J. V., Virbasius, C. A. and Scarpulla, R. C. (1993) *Genes Dev.* **7**, 380–392
- Evans, M. J. and Scarpulla, R. C. (1989) *J. Biol. Chem.* **264**, 14361–14368
- Virbasius, J. V. and Scarpulla, R. C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1309–1313
- Gugneja, S., Virbasius, J. V. and Scarpulla, R. C. (1995) *Mol. Cell. Biol.* **15**, 102–111
- Wasylc, B., Hanh, S. L. and Giovane, A. (1993) *Eur. J. Biochem.* **211**, 7–18
- Forsburg, S. L. and Guarente, L. (1989) *Annu. Rev. Cell Biol.* **5**, 153–180
- Ohta, S., Tomura, H., Matsuda, K. and Kagawa, Y. (1988) *J. Biol. Chem.* **263**, 11257–11262
- Neckelmann, N., Warner, C. K., Chung, A. et al. (1989) *Genomics* **5**, 829–843
- Peña, P. and Garesse, R. (1993) *Biochem. Biophys. Res. Commun.* **195**, 785–791
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Bankier, A. T., Weston, K. M. and Barrell, B. G. (1987) *Methods Enzymol.* **155**, 51–93
- Staden, R. (1986) *Nucleic Acids Res.* **14**, 217–231
- Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Pardue, M. L. (1986) *Drosophila, A Practical Approach*, pp. 111–137, IRL Press, Oxford
- Bucher, P. (1990) *J. Mol. Biol.* **212**, 563–578
- Cherbas, L. and Cherbas, P. (1993) *Insect Biochem. Mol. Biol.* **23**, 81–90
- Olson, E. N. (1990) *Genes Dev.* **4**, 1454–1461
- Kaushal, S., Schneider, J. W., Nadal-Ginard, B. and Mahdavi, V. (1994) *Science* **266**, 1236–1240

- 29 Chung, A. B., Stepien, G., Haraguchi, Y., Li, K. and Wallace, D. C. (1992) *J. Biol. Chem.* **267**, 21154–21161
- 30 Suzuki, H., Hosokawa, Y., Nishikimi, M. and Ozawa, T. (1991) *J. Biol. Chem.* **266**, 2333–2338
- 31 Villena, J. A., Martin, I., Viñas, O., Cormand, B., Iglesias, R., Mampel, T., Giral, M. and Villaroya, F. (1994) *J. Biol. Chem.* **269**, 32649–32654
- 32 Efiok, B. J. S., Chiorini, J. A. and Safer, B. (1994) *J. Biol. Chem.* **269**, 18921–18930
- 33 Höög, C., Calzone, F. J., Cutting, A. E., Britten, R. J. and Davidson, E. H. (1991) *Development* **112**, 351–364
- 34 Kassis, J. A., Poole, S. J., Wright, D. K. and O'Farrell, P. H. (1986) *EMBO J.* **5**, 3583–3589
- 35 Pan, D., Valentine, S. A. and Courey, A. J. (1994) *Mech. Dev.* **46**, 41–53
- 36 Maier, D., Preiss, A. and Powell, J. R. (1990) *EMBO J.* **9**, 3957–3966
- 37 Wilde, C. D. and Akam, M. (1987) *EMBO J.* **6**, 1393–1401
- 38 Beverley, S. M. and Wilson, A. C. (1984) *J. Mol. Evol.* **21**, 1–13
- 39 Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O. and Fields, C. (1992) *Nucleic Acids Res.* **20**, 4255–4262
- 40 Bowman, E. and Knock, T. E. (1992) *Gene* **114**, 157–163
- 41 Boutry, M. and Chua, N. (1985) *EMBO J.* **4**, 2159–2165
- 42 Takeda, M., Vassarotti, A. and Douglas, M. G. (1985) *J. Biol. Chem.* **260**, 15458–15465
- 43 Hendrick, J. P., Hodges, P. E. and Rosenberg, L. E. (1989) *Proc. Natl. Acad. Sci. U.S.A.*, 4056–4060
- 44 Kozac, M. (1991) *J. Cell. Biol.* **115**, 887–903
- 45 Kozac, M. (1994) *J. Mol. Biol.* **235**, 95–110
- 46 DeSimone, S. M. and White, K. (1993) *Mol. Cell. Biol.* **13**, 3641–3649
- 47 De la Pompa, J. L., Garcia, J. R. and Ferrús, A. (1989) *Dev. Biol.* **131**, 439–454
- 48 Fleming, R. J., DeSimone, S. M. and White, K. (1989) *Mol. Cell. Biol.* **9**, 719–725
- 49 Merika, M. and Orkin, S. T. (1993) *Mol. Cell. Biol.* **13**, 3999–4010
- 50 Ko, L. J. and Engel, J. M. (1993) *Mol. Cell. Biol.* **13**, 4011–4022
- 51 Soeller, W. C., Oh, C. E. and Kornberg, T. B. (1993) *Mol. Cell. Biol.* **13**, 7961–7970
- 52 O'Donnell, K. H. and Wensink, P. C. (1994) *Nucleic Acids Res.* **22**, 4712–4718
- 53 Pikó, L. and Taylor, K. D. (1987) *Dev. Biol.* **123**, 364–374
- 54 Elliot, D. and Jacobs, H. T. (1989) *Mol. Cell. Biol.* **9**, 1069–1082
- 55 El Meziane, A., Callen, J. C. and Mounolou, J. C. (1989) *EMBO J.* **8**, 1649–1655
- 56 Gangfeng, W. and Shaoyi, Y. (1992) *Int. J. Dev. Biol.* **36**, 477–482
- 57 Rubenstein, J. L. R., Brutlag, D. and Clayton, D. (1977) *Cell* **12**, 471–482
- 58 de Bruijn, M. H. L. (1983) *Nature (London)* **304**, 234–241