Volume 9 Number 18 1981 Nucleic Acids Research

Vitellogenin in Drosophila melanogaster: sequence of the yolk protein I gene and its flanking regions

Bernd Hovemann, Ricardo Galler, Uwe Walldorf, Hans Küpper[+] and Ekkehard K.F.Bautz

Institut für Molekulare Genetik, Universität Heidelberg, D-6900 Heidelberg, GFR

Received 26 June 1981

ABSTRACT

We have isolated recombinant DNA clones coding for female specific proteins from *Drosophila melanogaster*. By screening with $^{32}P-(A)^{+}RNA$ from male and female flies, respectively, we were able to isolate a set of 100 cDNA clones which showed a positive hybridization signal for RNA from female flies. These clones have been rescreened with RNA isolated from fat body of two day old male and female flies. We obtained four positive cDNA clones. Isolation of the corresponding genomic sequences, construction of the physical map and comparing it with the restriction maps published by Barnett et al. (1) led us to conclude that we had isolated the genes coding for two of the three known yolk protein precursors (vitellogenins), YP I and YP II. The sequence of the YP I gene was determined. It gives rise to a protein of 48 700 dalton MW which might be cleaved to a MW of 46 700 during transport. The coding sequence is interrupted by a single intron of 75 bases in length. The proposed leader sequence starts at a region homologous to six heat shock gene sequences at the site of initiation of transcription, suggesting the existence of an 11 bp cap specific consensus sequence for *Drosophila melanogaster*.

INTRODUCTION

During oogenesis the synthesis of many specific proteins is under hormonal control (2). Under the influence of ecdysone the fat body cells produce large amounts of yolk protein precursors, the vitellogenin I (46 000 daltons), II (45 000 daltons) and III (44 000 daltons) and secrete them into the hemolymph from which they are taken up by the developing oocyte to form yolk. Vitellogenin is also produced in the ovary itself where it may be induced by juvenile hormone (3). The process of yolk formation is accomplished by a two step process, a terminal cleavage of a signal peptide and a modification step (4). Recently (1) it was shown that the genes coding for the yolk protein in *Drosophila melanogaster* are single copy and dispersed at two chromosomal loci, YP I and II at 8F-9A, YP III at 12B-C.

It is obvious that egg formation represents an interesting system for the study of hormone induced protein synthesis, transport and modification. Here

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

we present the DNA sequence of the YP I gene region. In addition the protein sequence of YP I deduced from the DNA sequence data is presented.

MATERIALS AND METHODS

DNA and Enzymes

pBR 322 or 325 DNA was used as a cloning vector to transform *E. coli* HB 101. Reverse transcriptase of avian myeloblastosis virus was a gift of Dr. J.W.Beard, the Office of Program Resources and Logistics, NCI, USA. All other enzymes were from commercial sources. Plasmid DNA was isolated by a modification of the method of Clarke and Carbon (5).

Isolation of poly $(A)^+$RNA

Two to three day old flies were collected with a vacuum cleaner and frozen in liquid nitrogen. For the isolation of male and female specific RNA, etherized male and female flies were separated by hand before freezing. Frozen flies were powderized in a mortar cooled by liquid nitrogen. To 10 g of powder 50 ml of RNA buffer (20 mM Tris/Cl, pH 7.5, 100 mM NaCl, 0.5 % SDS, 1 mM EDTA, 75 µg/ml Heparin, 20 µg/ml polyvinylsulfate) and 50 ml of phenol.chloroform-isoamylalcohol (24:24:2) were added and the material was further homogenized in an omnimix (Sorvall) for 3' in an ice bath; after centrifugation at 10.000 rpm for 15' the aqueous phase was decanted and the volominous interphase was reextracted with 1/2 vol of RNA buffer. After centrifugation the supernatants were combined and reextracted with phenol:chloroform-isoamylalcohol. The RNA was precipitated from the aqueous phase and adjusted to 0.3 M NaOAc, pH 5. Poly $(A)^+$RNA was prepared by two cycles of oligo (dT)-cellulose chromatography.

Poly $(A)^+$RNA was further purified by sucrose gradient centrifugation (containing 10 mM Tris/Cl pH 7.5, 0.5 mM EDTA, 0.1 % SDS and from 10 to 30 % sucrose) for 16 hrs at 35 000 rpm in a SW 40 rotor at room temperature. Fractions were tested in a reticulocyte cell free translation system and the fractions directing the synthesis of polypeptides in the size range of 20 to 80 kd were used for cloning.

Synthesis of ds-cDNA

cDNA was synthesized from sucrose gradient-enriched RNA. In a 250 µl reaction (50 mM Tris/Cl, pH 8.3, 10 mM Mg $(OAc)_2$, 100 mM KCl, 10 mM DTT, 15 µg oligo-dT_{12-18}, 1 mM each of dATP, dGTP, dCTP and dTTP and all four α-P-32-deoxynucleosidetriphosphates (total of 40 µCi), 25 µg poly $(A)^+$ RNA were incubated with 60 units of AMV-reverse transcriptase at 42^oC for 45 min. 250 µl of dilution buffer was added (5mM DTT, 5mM Tris/Cl, pH 8.3) and the mixture

was incubated for another 50 min at 45 °C. After boiling for 2 min, the reaction mixture was chilled quickly. To render the DNA double stranded, an extra 100 mM Hepes buffer (pH 6.9), 100 mM KCl and 100 μM of each of the dXTPs were added and incubation was continued at 15 °C for 4 hrs with 20 units of E. coli DNA pol. I. The reaction was stopped by the addition of EDTA and the mixture was passed through a Sephadex G 50 column to remove the radioactive precursors. The double stranded cDNA was precipitated with EtOH and digested with nuclease S_1 to open the hairpin structure. After extraction with phenol and precipitation with EtOH the cDNA was used for insertion into pBR 322.

Cloning of ds-cDNA

Oligo dC-tails were added to the cDNA by incubation with terminal transferase. This cDNA was hybridized to Pst I - opened, oligo-dG-tailed pBR 322 DNA and hybrid molecules were transformed into E. coli HB 101. Colonies were selected for Tet^R and Amp^S phenotype.

Screening of colonies for female specific sequences

Tet^R and Amp^S colonies were transferred to microtiter plates and nitrocellulose filters carrying lysed bacterial colonies were prepared according to Grunstein and Hogness (6). Duplicates of the filters were hybridized to male specific and female specific $poly(A)^+$ RNA which was labeled with P-32 by incubation at pH 9.5 at 90 °C for 75 min and subsequent incubation with T_4 polynucleotide kinase and γ-^{32}P-ATP at 37 °C for 30 min (specific activity: 2×10^7 cpm/ug RNA). One hundred colonies which showed a difference in hybridization intensity were kept on two filters for rescreening with cDNA complementary to male and female specific RNA from fat body.

Isolation of genomic clones

For the isolation of genomic clones, the Drosophila melanogaster (Canton S) library from T. Maniatis was used. All screening procedures were as described by Maniatis et al. (7). The isolated HindIII fragments containing either one of the two isolated YP genes were recloned in pBR 322 or pBR 325.

Restriction mapping

Cloned cDNA and isolated λ phage DNA were mapped by the double digestion method or by the method of Smith and Birnstiel (8). Restriction fragments were separated either on 0.5-1.5 % Agarose gels or on 3-10 % Polyacrylamide gradient gels. As molecular weight standards HindIII digested λ phage and HinfI digested pBR 322 DNA were used. DNA was transferred from Agarose slab gels to nitrocellulose filters by a modification of the method of Southern (9). Agarose gels were pretreated with 0.25 M HCl to reduce the time of transfer to 3 hrs. Hybridization of ^{32}P-DNA probes to filter bound DNA was usually done for

1 day with a concentration of 1 µg labelled probe/100 ml hybridization solution.

DNA sequence analysis

DNA sequencing was performed according to Maxam and Gilbert (10). 6 % sequencing gels were prepared and dried down as described by Garoff et al. (11). Pst I cut fragments were 3' end labelled with α-^{32}P-cordycepintriphosphate (12), all other sequencing was from 5' end labelled fragments using T_4 polynucleotide kinase. To obtain clean fragments for sequencing the labelled DNA was routinely run through two gels, one after kinasing and one to separate the one end labelled fragments after second restriction cleavage.

S-1 analysis

Five µg poly(A)$^+$ RNA and about 0.5 µg labelled DNA fragment were precipitated with 20 µg of carrier tRNA, pelleted and taken up in 30 µl of S-1 hybridization buffer (80 % Formamide, 0.4 M NaCl, 0.04 M Pipes pH 6.4, 0.001 M EDTA) (13). After heating at 80 ^{o}C for 5 min the solution was kept at 60 ^{o}C for 3 hrs. Then 10 volumes of S-1 digestion buffer was added (0.25 M NaCl, 0.03 M NaOAc, 0.001 M $ZnSO_4$, 5 % glycerol) containing 1000 units of nuclease S-1 (Boehringer, Mannheim) and incubation was continued for an hour at 42°C. After precipitation with 10 µg of carrier tRNA the sample was run through a 6 % polyacrylamide-sequencing gel.

RESULTS

Construction of cDNA clones from female specific message

In order to study genes whose expression is regulated by hormones we attempted to clone DNA sequences complementary to female specific messages. This was done by preparing cDNA from poly(A)$^+$ RNA of 2-3 day old flies and subsequent cloning of this cDNA into the PstI site of pBR 322 via oligo dC/dG tailing. We picked 850 Amp^S, Tet^R colonies and used them in the screening procedure with poly(A)$^+$ RNA prepared from 2-3 day old male and female flies. Finally a collection of 100 putative female specific clones was selected and kept on nitrocellulose filters for further screening with tissue specific probes such as RNA from fat body or ovaries.

Screening of cDNA collection and phage library

Screening with cDNA to poly(A)$^+$ RNA isolated from fat body of 2-3 day old female and male flies, respectively, gave a positive signal with four of the cDNA clones*). The positive clones appear to derive from two classes of RNA

*) We thank Dr. Bownes, University of Edinburgh, for performing the screening with poly(A)$^+$ RNA isolated from fat bodies of Drosophila melanogaster.

since two clones each showed overlaps in their restriction maps. From each pair of cDNA clones we used the one with the longest inserted cDNA (cDm-42 and cDm-46, respectively) both for screening the library of genomic DNA and for in situ hybridization to polytene chromosomes.

cDm-42 and cDm-46 DNA were labeled with (^{3}H) dTTP by nick-translation (14) to a specific activity of 5 x 10^6 cpm/ug and hybridized, in situ to polytene chromosomes (data not shown). Both clones hybridized to a only one chromosomal region which coincides with 8E/9A, the known loci for YP I and YP II.

In order to screen the genomic library cDm-42 and -46 DNA was nick-translated with α-^{32}P-dCTP to a specific activity of 0.5 - 1.0 x 10^8 cpm/ug. Twenty positive plaques were obtained, of which 3 phages (DmF$_3$ A-C) containing segments of the same chromosomal region were selected for further analysis. The locations of the overlapping segments with respect to the total inserts are shown in Fig. 1.

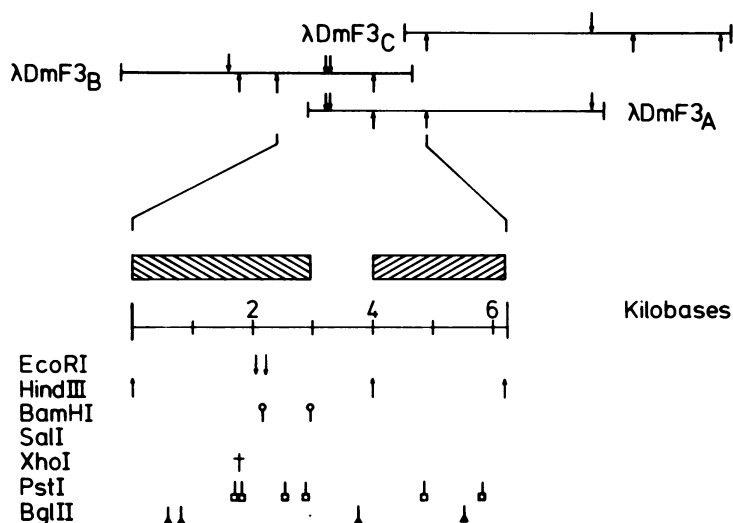

FIG. 1

Comparison of lambda clones carrying segments which gave positive Southern blots with female specific cDM 42 and -46 DNA. Each probe gave a positive hybridization signal at one region and a weaker one at a different site separated by a fragment which did not hybridize to either one of the hybridization probes (hatched boxes) indicating two related genes. The physical map of restriction enzymes recognizing hexamer sequences confirmed the assumption that the cloned Drosophila melanogaster DNA contained the genes for YP I and YP II also cloned by Barnett et al. (1) and Riddell et al. (15).

Restriction analysis of genomic clones

Restriction analysis by double digestion with EcoRI and HindIII revealed the cleavage pattern known from the work of Barnett et al. (1) to be the coding region for the YP I and YP II genes from *Drosophila melanogaster*. Further analysis of the two HindIII fragments, positive in Southern probing, with other restriction enzymes recognizing hexamer sequences and subsequently probing the blotted agarose gel with cDm-42 and -46 DNA (lower portion of Fig. 1) confirmed the assumption that we had isolated the genes coding for YP I and YP II.

The restriction map obtained is consistent with the one presented by Barnett et al. (1) and Riddell et al. (15) except for an additional PstI fragment of 0.1 kb in the 3' terminal part of the YP I gene sequence. The cDNA probes which cover about one kb of the 3' terminal mRNA sequence show remarkable crosshybridization even under stringent conditions (data not shown). Both genes have been recloned as HindIII fragments in pBR 325 (YP I) and pBR 322 (YP II) (Fig. 2).

Sequencing strategy and analysis of the YP I gene

In order to determine the actual coding region we first sequenced a portion of the cDNA clone starting from the poly A tail towards the coding region (Fig. 3). The corresponding genomic DNA was then sequenced beginning with the poly A addition site. The actual coding sequence was determined by S1 digestion of labelled DNA fragments after hybridization with poly $(A)^+$ RNA from female flies (Fig. 4).

The following features are evident from the sequence (Fig. 3). A characteristic 'TATA' box is present 23 bp upstream of the putative cap-site which closely matches in sequence with the transcrpiptional start of six heat shock gene messages from *Drosophila melanogaster.* This extensive homology strongly suggested the location of the 5' end of mRNA, the cap site, at this place, possibly at position 1 (Fig. 3, pos. 3-14, underlined). Also transcription experiments obtained with a whole cell extract from HeLa cells lead to run-off products of about 520 NT for a Bst E I cut template (pos. 516, unpublished results). The sequence AATAAA preceding most poly(A) addition sites was found at position 1606 with poly A added at position 1634 or 1635.

It is interesting to note that all 10 Eco RII restriction sites showed a missing second "C" of the CC$^{A}_{T}$GG recognition site in the sequencing ladder due to methylation as described by Ohmori et al. (16). Also the Taq I site in position 756 (Fig. 3) was not cleaved in mapping experiments which could be explained according to Razin et al. (17) who showed that GATC is the site of methylation by the "dam" protein. The possibility that Taq I does not cleave

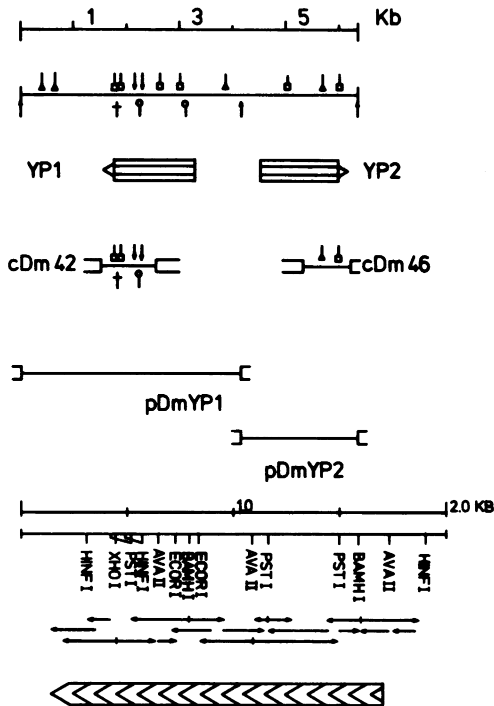

FIG. 2

Restriction map of two HindIII fragments carrying YP I and II

The restriction map for EcoRI - , HindIII - , BamHI - , PstI - , XHOI - and BglIII - is shown together with the YP I and YP II coding region (1, 15) indicated by arrows pointing towards the direction of transcription. The physical map of the cDNA clones Dm 42 and 46 is aligned to the corresponding genomic sequences. Both coding regions have been recloned as HindIII fragments. The lower part shows a more detailed restriction map of that portion of pDm YP I which has been sequenced. This is equivalent to the region occupied by the leftward pointing arrow in the upper part of the figure. Single arrows below the detailed restriction map indicate regions covered by sequencing.

when its recognition site is superimposed by a Sau 3A site carrying a methylated "A" was already discussed by Streeck et al. (18).

The YP I Protein

The coding sequence starting at the first AUG (Pos. 62) downstream from the cap site runs into a cluster of 4 stop codons around nucleotide 290 which terminate in all three reading frames. The interruption of translation is followed by a very AT rich DNA sequence including an unusual stretch of A_{16}.

```
        YP 1
        10         20         30         40         50         60         70         80         90         100
-100 ..........  ....CCTGAG CCAGCGAAAA GCAAGTCGGA AAATGGGAAA TCGCTCAGCG TAAATTGTGG TATATAAACC ACCATCGTTG GATTTGGAAG
   0 GCCAGTTCAA CTCACTCAGT GTTGAAGTCG CATCCGCAGG ACCAAATCCC AAATCCGAAC CATGAACCCC ATGAGAGTGC TGAGCCTTCT GGCTTGCTTG
 100 GCGGTCGCCG CCTTGGCCAA GCCCAATGGC CGTATGGACA ACTCCGTCAA CCAGGCATTG AAGCCGTCGC AGTGGCTCTC CGGATCCCAG CTGGAGGCCA
 200 TTCCCGCCCT CGACGATTTC ACCATTGAGC GTCTGGAGAA CATGAACCTG GAGCGTGGCG CCGAGCTGCT GCAGCAAGTC TGTGAGTAAT CCTAGATGCA
 300 GATAAAAAAA AAAAAAAAAC CTCGAATATT CTATGGAATA TATATATCCT TTGTAGACCA CCTGTCGCAG ATCCACCACA ACGTTGAGCC CAACTATGTG
 400 CCCAGCGGCA TCCAGGTCTA TGTGCCCAAG CCCAATGGTG ACAAGACCGT TGCTCCCCTG AACGAGATGA TCCAGCGCCT GAAGCAGAAG CAGAACTTTG
 500 GTGAGGATGA GGTGACCATC ATTGTGACCG GACTGCCCCA GACCAGCGAG ACCGTGAAGA AGGCGACCAG GAAGCTGGTT CAGGCTTACA TGCAGCGCTA
 600 CAATCTGCAG CAGCAGCGCC AGCACGGCAA GAACGGCAAC CAGGACTACC AGGATCAGAG CAACGAACAG AGGAAGAACC AGAGGACCAG CAGCGAGGAG
 700 GACTACAGCG AGGAGGTTAA GAACGCCAAG ACCCAAAGCG GCGACATCAT TGTGATCGAT TTGGGCTCCA AGCTGAACAC CTATGAGCGT TATGCCATGC
 800 TCGACATTGA GAAGACCGGC GCCAAGATCG GCAAGTGGAT CGTCCAGATG GTCAACGAGT TGGACATGCC CTTCGATACC ATTCACCTGA TTGGCCAGAA
 900 TGTGGGTGCC CATGTTGCCG GTGCCGCTGC CCAGGAATTC ACCCGTCTCA CCGGACACAA GCTGCGCCGT GTCACCGGTC TGGATCCCTC CAAGATCGTG
1000 GCCAAGAGCA AGAACACCCT GACCGGTCTG GCTCGCGGTG ATGCTGAATT CGTTGACGCC ATCCACACCT CGGTCTACGG CATGGGCACC CCCATCCGCT
1100 CCGGTGATGT TGACTTCTAT CCCAATGGAC CTGCCGCCGG TGTTCCCGGA GCCAGCAACG TGGTGGAGGC CGCCATGCGT GCCACCCGCT ACTTCGCCGA
1200 GTCCGTGCGT CCCGGAAACG AGAGGAGCTT CCCCGCCGTG CCAGCCAACT CCCTGCAGCA GTACAAGCAG AACGATGGAT TCGGCAAGCG TGCCTACATG
1300 GGCATCGATA CCGCTCACGA TCTCGAGGGT GACTACATTC TGCAGGTGAA CCCCAAGTCT CCTTTCGGCC GCAACGCACC CGCCCAGAAG CAGAGCAGCT
1400 ACCACGGTGT CCACCAGGCG TGGAACACCA ACCAGGACAG CAAGGACTAC CAGTAAGGAT GAGTCTGCTT ACTCTGGACA CCTGGAATGG CAACTACCAA
1500 ACAACCACCC AACCACACAA ACACTGTAGT CCCTAAGTTG AACCCATATT GGCCCTTTTC TTGAGATTAC CTAAACATTT AACGAGCACA TCGCGAAATT
1600 CAGCAAATAA ACGCTCGATA AAGAGCTTAA AAATACC

        YP1.CDNA

     CAGCAAATAA ACGCTCGATA AAGAGCTTAA AAATAAAAAAAAAAAACCCCCCCCCCC...PBR322
```

S-1 mapping experiments (Fig. 4) revealed the existence of an intron of about 75 NT in length starting and ending approximately at positions 282 and 356 respectively. At these positions we find putative intron boundaries (Fig. 3). Excluding the intron a coding sequence of 1317 nucleotides starting at position 62 and terminating at position 1454 is found (Fig. 3 corresponding to pos. 62 and pos. 1379 in Fig. 5). It gives rise to a protein of 48 700 dalton MW. The other two reading frames are unusable for translation since they contain a total of 30 stop codons (Fig. 5, triple stars). The NH_2-terminal protein extension shows an amino acid sequence of extremely hydrophobic character as expected for amino acids comprising a signal peptide. Substraction of this putative signal sequence results in a mature protein of 46 780 dalton MW. This is in agreement with the known size of the YP I protein from SDS gel electrophoresis.

DISCUSSION

Vitellogenin from *Drosophila melanogaster* has been studied for several years but only recently it was recognized to consist of three different polypeptides which are not derived from a unique gene. Final proof came from the work of Barnett et al. (4) who described three different coding regions, two clustered at the chromosomal locus 8F/9A and a third one about 1 000 kb away at the locus 12D, all of which appear to be transcribed and translated with equal efficiency. The *in vivo* expression of vitellogenin as an egg storage protein is sex and time specific. Due to their similarity it is very difficult to separate and purify the three yolk proteins; consequently the protein sequences are not available. We isolated two of the three YP genes, YP I and YP II. Three genomic clones carrying two coding sequences or part of it were isolated from the lambda library of genomic (Canton S) DNA. They were identified by their chromosomal location and their restriction patterns which are identical to the ones determined by Barnett et al. (1) and Riddell et al. (15). As a first

FIG. 3
Complete nucleotide sequence of the YP I gene
The sequence, noncoding strand only, shows the 5' flanking region including a "TATA" box. The proposed site of initiation of transcription, pos. 1, the intron splice points pos. 281 and 356, and the polyadenylation site, pos. 1634 or 1635 are indicated by arrows. The first ATG, pos. 62, underlined, following the proposed site of initiation of transcription starts an open reading frame of, excluding the intron, 1317 nucleotides. The stop codon, pos. 1454, and the polyadenylation signal are also underlined. The lower part shows a sequence from the cloned cDNA, cDm 42 suggesting the site of polyadenylation to be either T or A.

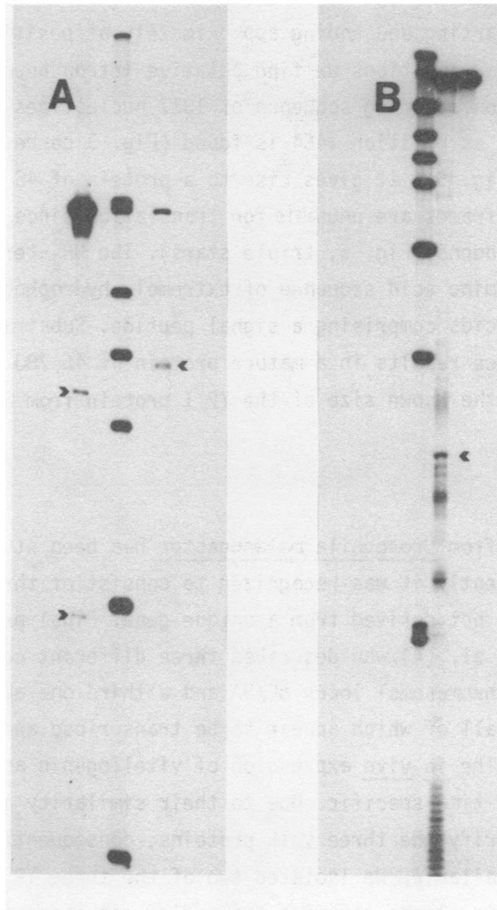

FIG. 4

Intron boundary sites determined by nuclease S1 digestion

A: The AvaII - BamHI restriction Fragment, fig. 3, pos. 181 to 684 5' end labelled at the AvaII site, was hybridized with poly(A)$^+$ RNA from female flies and treated with nuclease S2. The left lane shows the untreated fragment and the right lane the S1 digestion product of 331 nucleotides in length and some undigested material. In the middle lane HinfI digested pBR 322 DNA fragments were run as molecular weight markers. The two minor bands in the left lane (arrows) are probably due to single strand cleavage by BamHI at the sequence positions 373 and 472, giving rise to two new bands of 314 and 215 nucleotides under denaturing conditions only.
B: The 800 nucleotides BamHI fragment, fig. 3, pos. 181-981, was 3' end labelled with pol. I and α-^{32}P-triphosphates to determine the left (5'-)end of the intron. After S1 digestion a series of bands appeared. As shown in the middle lane a strong band on top (arrow), 95 nucleotides in length, results in a splice point at position 281. The additional bands may be due to crosshybridization to homologous YP II sequences. The right lane shows undigested BamHI fragment, the left lane HinfI digested pBR 322 DNA and an additional marker fragment of 900 nucleotides. Both S1 experiments are analyzed on 6 % "sequencing" gels.

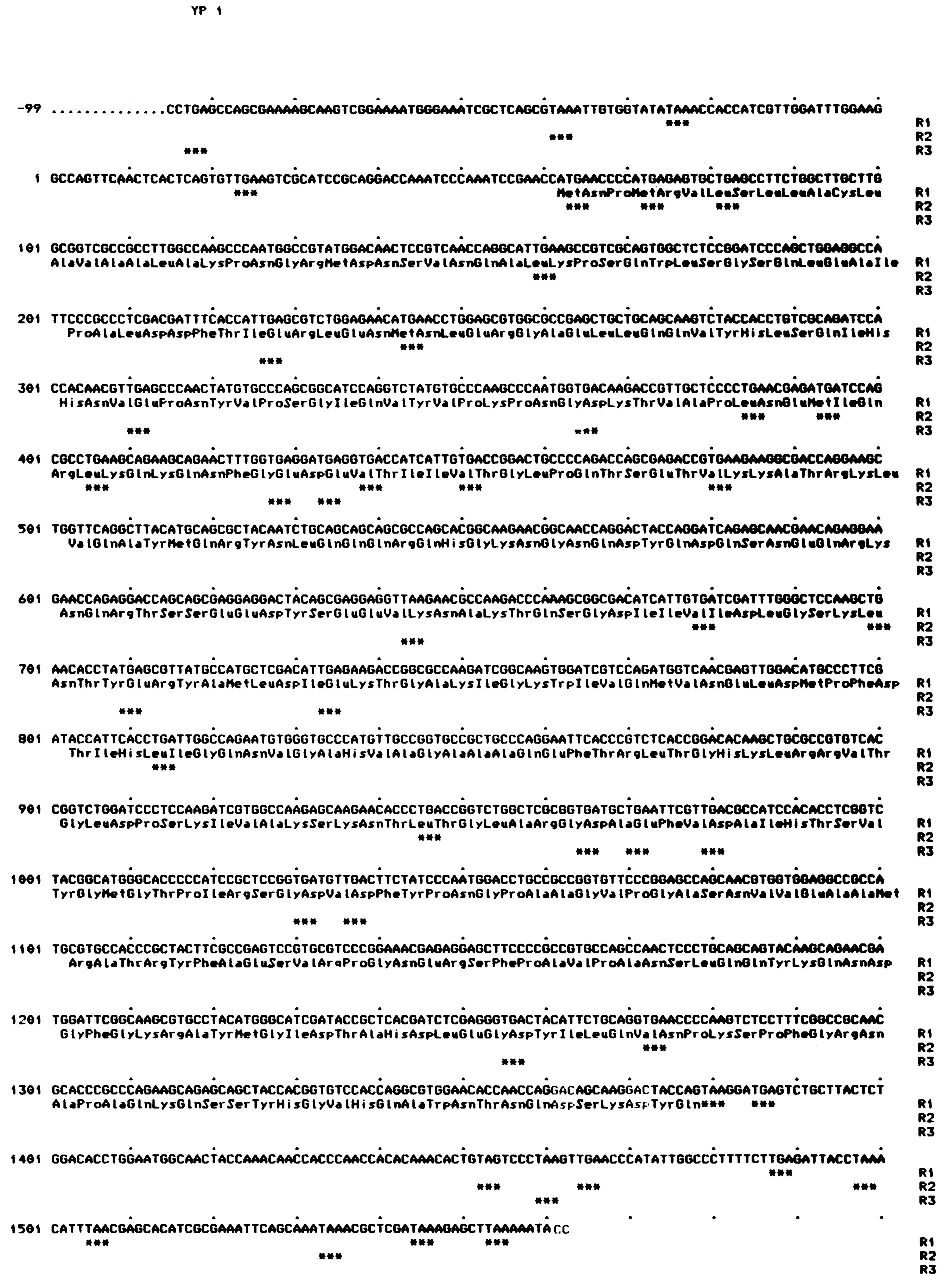

FIG. 5

Complete nucleotide sequence of the YP I coding region (excluding the putative intron) and its amino acid sequence

The termination codons are indicated by triple asterisks.

step to study the expression of the YP genes on a molecular basis we determined the primary sequence of the YP I gene and its 5' flanking sequence:

Although the region transcribed as YP I mRNA was known from S-1 mapping experiments (4, 15) the exact start of the mRNA transcript is still tentative. A characteristic "TATA" box, however, is evident 23 bases upstream from the proposed 5' end of the mRNA (pos. 1 in fig. 3). The TATA sequence is framed by GC-rich regions as found next to most published eucaryotic promoter or selector sites. Comparing six heat shock and the YP I gene sequence from *Drosophila melanogaster* a consensus capping sequence is found near the proposed site of initiation of transcription (the sequences are aligned to give maximum homology):

hsp 22 (23)	CAG......TTCA
hsp 23 (23)	CAG.TTGAATTCA
hsp 26 (23)	CAGATCGAATCCA
hsp 27 (23)	CAGTCTAAACTGA
hsp 70 clone G 13 (20)	CAATTCAA.TTCA
hsp 70 clone BHI (20)	CAATTCAA.TTCA
YP I	CAGTTCAA.CTCA
consensus sequence	CA$^{G}_{A}$.TCAA.$^{T}_{C}$TCA

No homology, however, can be recognized to a proposed conserved sequence 70 to 80 bp prior to transcription initiation as found for several other polymerase II transcribed genes (21). Perhaps there is an as yet unresolved recognition site which modulates gene expression in *Drosophila*.

The length of the mature YP I RNA, excluding the poly (A) tail of unknown length was found to be approximately 1560 nucleotides. Its primary transcript is interrupted by at least one intron. The 5' and 3' boundary sequences of the YP I intron share homology with the flanking sequences from two *Drosophila melanogaster* actin gene introns (24):

Dm A_4 (24)	GTGCGTGG.........TATCCTGCAG
Dm A_6 (24)	GTGCGTGA......TGTCCTGTTCAG
YP I	GTGAGTAA......TATCCTTTGTAG

Other small introns cannot be excluded as long as the mRNA sequence itself is unresolved, but it seems unlikely in view of the high number of stop codons in the remaining two reading frames.

Although the mechanism of transcription termination and polyadenylation

is still unclear, one general feature of all eucaryotic mRNA 3' ends is a "AATAAA" sequence preceding the polyadenylation site by 10 to 30 bases. This turned out to be true also for the YP I sequence presented here.

The coding sequence of 1317 nucleotides includes codons for a signal peptide that facilitates the transport of the vitellogenin polypeptide. Only three out of the first 19 amino acids are not hydrophobic. Its leucine content is rather high which is characteristic for signal sequences. Since there are no published results on the N-terminal amino acid sequence of the mature protein, the suggested N-terminus of the mature protein at the end of the block of hycrophobic amino acids, i.e. at position of amino acid 20 (fig. 5) remains speculative.

Comparison of the cDNA and the corresponding genomic DNA sequence suggests the poly (A) addition site to be either nucleotide 1634 or 1635 (fig. 3).

We hope that knowing the sequence of the YP I gene will prove helpful in the purification of the YP I protein using antibodies produced against synthetic polypeptides (25) and subsequently in the verification of the polypeptide sequence by partial amino acid sequencing.

ACKNOWLEDGEMENT

We are grateful fo Drs. C. Flyzanis and W. Roewekamp for help in the initial screening of the lambda library. We thank R. Kabisch for *in situ* hybridization experiments, Drs. H. Weiher and E. Beck for their support during sequencing, N. Hernandes and Dr. W. Keller for a gift of HeLa cell extract and helpful discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (Forschergruppe Genexpression) and by a predoctoral fellowship from the National Research Council of Brazil (C.N.Pq) to R. Galler.

[+]Present address: Biogen, S.A., Geneva, Switzerland

REFERENCES

1 Barnett, T., Pachl, G., Gergen, J.P. and Wensink, P.C. (1980) Cell 21, 729-738.
2 Wyatt, G.R., and Pan, M.L. (1978) Ann. Rev. Biochem. 47, 779-817.
3 Postlethwait, J.H., Bownes, M., and Jowett, T. (1980) Dev. Biol. 79, 379-387.
4 Warren, T.G., Brennan, M.D., and Mahowald, A.P. (1979) Proc. Natl. Acad. Sci. USA 76, 2848-2852.
5 Clarke, L., and Carbon, J. (1975) Proc. Natl. Acad. Sci. USA 72, 4361-4365.
6 Grunstein, M., and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.

7 Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C. Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell 15, 687-701.
8 Smith, H.O., and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398.
9 Southern, E.M. (1975) J. Mol. Biol. 93, 503-517.
10 Maxam, A.M., and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
11 Garoff, H., and Ansorge, W. (1981) Analytical Biochemistry, in press.
12 Tu, C.-P.D., and Cohen, S.N. (1980) Gene 10, 177-183.
13 Sharp, P.A., Berg, A.J., and Buget, S.M. (1980) Methods in Enzymol. 65, 750-768.
14 Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
15 Riddell, D. Chl., Higgins, M.J., McMillan, B.J., and White, B.N. (1981) Nucl. Acids Res. 9, 1323-1388.
16 Ohmori, H., Tomizawa, J., and Maxam, A.M. (1978) Nucl. Acids Res. 5, 1479-1485.
17 Razin, A., Urieli, S., Pollack, Y., Gruenbaum, Y., and Glaser, G. (1980) Nucl. Acids Res. 8, 1783-1792.
18 Streeck, R.E. (1980) Gene 12, 267-275.
19 Goldberg, D. (1979) Ph. D.thesis, Stanford University.
20 Ingolia, T.D., Craig, E.A., and McCarthy, B.J. (1980) Cell 21, 669-679.
21 Benoist, C., O'Hara, K., Breathnach,R., Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
22 Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
23 Ingolia, T.D., and Craig, E.A. (1981) Nucl . Acids Res. 9, 1627-1942.
24 Fryberg, E.A., Beverly, J.B., Hershey, N.D., Mixter, K.S., and Davidson, N. (1981) 24, 107-116.
25 Walter, G., Scheidtmann, K.-H., Carbone, A., Laudano, A.P. and Doolittle, R.F. (1980) Proc. Natl. Acad. Sci. USA 77, 5197-5200.