Volume 15 Number 3 1987 Nucleic Acids Research

Structural analysis of the *Drosophila* rpA1 gene, a member of the eucaryotic 'A' type ribosomal protein family

Su Qian, Jing-Yu Zhang[1], Mark A.Kay and Marcelo Jacobs-Lorena*

Department of Developmental Genetics and Anatomy, School of Medicine, Case Western Reserve University, 2119 Abington Road, Cleveland, OH 44106, USA

Received October 7, 1986; Revised and Accepted January 9, 1987

ABSTRACT

The expression of ribosomal protein (r-protein) genes is uniquely regulated at the translational level during early development of Drosophila. Here we report results of a detailed analysis of the r-protein rpA1 gene. A cloned DNA sequence coding for rpA1 has been identified by hybrid-selected translation and amino acid composition analysis. The rpA1 gene was localized to polytene chromosome band 53CD. The nucleotide sequence of the rpA1 gene and its cDNA have been determined. rpA1 is a single copy gene and sequence comparison between the gene and its cDNA indicates that this r-protein gene is intronless. Allelic restriction site polymorphisms outsite of the gene were observed, while the coding sequence is well conserved between two Drosophila strains. The protein has unusual domains rich in Ala and charged residues. The rpA1 is homologous to the "A" family of eucaryotic acidic r-proteins which are known to play a key role in the initiation and elongation steps of protein synthesis.

INTRODUCTION

The coordinate synthesis of the many protein and RNA molecules that make up the ribosome must involve a complex series of regulatory mechanisms that are well documented but poorly understood. Ribosome synthesis requires over 70 genes and three different RNA polymerases. The rate of ribosome synthesis may vary dramatically depending on the developmental stage or the growth state of the cell. Under these conditions of varying demand, it is unclear how the different genes are coordinately up or down regulated. There is substantial evidence to suggest that in bacteria as well as in higher organisms, regulation occurs at a variety of levels of gene expression including transcription, RNA processing, translation and protein turnover (1-4). With the possible exception of autogenous translational regulation in bacteria (4), factors that mediate regulation of ribosomal protein (r-protein) gene expression have not been identified. The identification of such factors represents an important goal for future research and requires the detailed characterization of the genes involved.

The gene coding for Drosophila r-protein rpA1 has been first identified

in view of its unique pattern of expression during early Drosophila development (5). While most abundant mRNAs are continuously translated through all of oogenesis and embryogenesis, the translation of rpA1 and most other r-protein mRNAs is regulated differently: r-protein mRNAs are polysome-associated during oogenesis, largely excluded from polysomes in early embryos (when no ribosomes are synthesized) and again polysome-associated in late embryos (3, 6). As a further step in the elucidation of this selective translational regulation we have analyzed in detail the structure of the rpA1 gene. We determined the complete nucleotide sequence of genomic and cDNA clones. The data show that this gene is intronless and exists as single copy in the genome. Although the gene is conserved between two Drosophila strains, several restriction site polymorphisms have been detected in the regions surrounding the gene. The rpA1 protein has an unique amino acid sequence that identifies it as a member of the "A" family of acidic eucaryotic r-proteins. These "A"-type proteins are related to the E. coli r-protein L7/L12 (7-10), exist as dimers at the tip of the large ribosomal subunit, and play a key role in the initiation and elongation steps of protein synthesis (7, 11-13).

MATERIALS AND METHODS

Isolation of rpA1 genomic and cDNA clones.

A genomic library prepared from DNA of the Drosophila melanogaster Canton S strain (14) was screened using as a probe the previously isolated rpA1 cDNA clone (3, 5). A positive phage was isolated, restriction mapped and determined to contain an insert approximately 20 kb long. Northern blot analysis indicated that rpA1 was the only abundant embryonic RNA coded by the insert. For further analysis, a 2.4 kb Bam H1 fragment containing the entire rpA1 gene was subcloned into pBR322 and named p5D. Full-length rpA1 cDNA clones were obtained by using p5D as a probe to screen a lambda gt10 cDNA library prepared from poly(A)-containing RNA from 3 to 12 h-old Drosophila embryos of the Oregon R strain (15).

Hybrid-selected translation.

Total and poly(A)-containing RNA were isolated from 2 to 18 h-old embryos as described (16). About 30 ug of plasmid DNA was denatured in 0.3 N NaOH for 10 min at 65 ^{o}C, cooled to 4 ^{o}C, neutralized with an equal volume of 2 M NH_4Ac and applied onto nitrocellulose filters 13 mm in diameter. The filters were baked for 2 h at 80 ^{o}C under vacum and prehybridized for 1 h at 50 ^{o}C with a buffer containing 65% formamide, 400 mM NaCl, and 10 mM PIPES, pH 6.4.

Usually 2 filters were hybridized for 12 to 16 h at 50 °C with 40 ug of embryonic poly(A)-containing RNA disolved in 200 ul of the same buffer. The filters were washed and the hybridized RNA was eluted (17). The selected rpA1 mRNA was translated in the presence of ^{35}S-methionine and the products were analyzed by two-dimensional gel electrophoresis using isoelectric focusing in the first dimension (5).

Nucleotide sequence analysis.

Appropriate restriction fragments from cloned genomic or complementary DNAs were subcloned into M13mp18, mp19 (18) or into pEMBL mp18, mp19 (19) and their sequence determined by the dideoxy chain termination method of Sanger et al. (20). When sequencing long fragments, synthetic oligomers were used to prime the reaction from internal points.

S1 nuclease protection and primer extension assays.

The strategy for determination of the 5' end by S1 nuclease mapping is illustrated at the bottom of Fig. 4. A Sal I fragment containing the 5' portion of the rpA1 gene (Bam HI-Sal I segment) and part of vector pBR322 sequence was isolated from p5D. The Sal I site was labeled with ^{32}P using T4 polynucleotide kinase (21). The end-labeled DNA was hybridized to embryonic poly(A)+ RNA, treated with S1 nuclease (21), and the size of the protected fragment was analyzed on a sequencing gel.

For primer extension, a synthetic oligomer (17mer) complementary to a region 46 bp upstream of the Sal I site was kinase labeled and used to prime a reverse transcription reaction. The expriment was performed according to Domdey et al. (22) except that the incubation was for 1 h at 37 °C. Ten micrograms of poly(A)+ RNA was used in each reaction. The same oligomer was used to prime a cDNA sequencing reaction 'G' (dideoxyguanosine reaction) and the product was fractionated in the same sequencing gel as the primer extension product.

Amino acid composition analysis.

Total r-protein was prepared as described by Kay and Jacobs-Lorena (3). To purify rpA1, r-proteins were fractionated by 2-dimensional gel electrophoresis using isolectric focusing in the first dimension and SDS-polyacrylamide gel electrophoresis in the second dimension (5). After brief staining with Coomassie Blue, the rpA1 spot was excised and the protein was electroeluted (23) with an ISCO model 1750 electroelution apparatus. The rpA1 eluate was dialyzed against 2 mM NaH_2PO_4, 0.05% SDS for 2 days and lyophilized. Usually 20 mg of r-proteins were fractionated on four 1.5 mm-thick gels yielding approximately 20 ug of rpA1. The protein was

hydrolyzed in 6 N HCl for 16 h and the amino acid composition was determined with a Beckman amino acid analyzer model 119CL.

Analysis of the rpA1 gene organization in fly DNA.

To prepare genomic DNA, adult flies were homogenized at 4 ^{o}C in homogenization buffer I (1 ml per 100 flies): 60 mM NaCl, 10 mM Tris-HCl (pH 7.5), 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine, 5% sucrose. The homogenate was tranfered to one volume of digestion solution: 0.1 M Tris-HCl (pH 9.5), 0.5 M EDTA, 1.0% SDS, 2 mg/ml pronase and the mixture was incubated at 65 ^{0}C for 1 h. Potassium acetate was added to the final concentration of 1 M and the mixture was cooled on ice for 45 min. After centrifugation for 10 min at 14.000xg to remove debris, the nucleic acids were precipitated with two volumes of ethanol. The pellet was dissolved in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and the nucleic acids were digested with 0.1 mg/ml of RNAase (DNAase-free) for 2 h at 37 ^{o}C. The DNA was phenol extracted, dialyzed against TE overnight and ethanol precipitated.

Single fly DNA was prepared by a procedure modified from K. Burtis (Stanford University). Individual females were homogenized in 50 ul of homogenization buffer II: 150 mM NaCl, 10 mM Tris-HCl, 10 mM EDTA, 0.2% NP-40, pH 8.0. The homogenate was microfuged (1-2 sec) through glass wool pluged in a 1-ml pipetor tip. The pipetor tip was placed in a 0.5 ml microfuge tube for the spin. Homogenization buffer II (2 X 100 ul) was used to rinse the homogenizer and then the glass wool by centrifugation. The combined filtrate was microfuged at 14,000Xg for 1 min. The nuclear pellet was resuspended in 20 ul of homogenization buffer II and mixed with 60 ul of lysis buffer: 300 mM NaCl, 50mM Tris-HCl, 10 mM EDTA, 1% sarkosyl, pH 8.0. The mixture was extracted once with phenol/chloroform and ethanol precipitated. The pellet was dissolved in 50 ul of TE containing 50 ug/ml of boiled RNAase, incubated for 30 min at 37^{o}C followed by phenol/chloroform extraction and ethanol precipitation.

Restriction digestion was carried out with 10 units of enzyme per microgram of DNA for 4 h at 37 ^{o}C and fractionated by electrophoresis on 0.8% agarose gels (single fly DNA was digested with 5 units of Rsa I and fractionated on a 2.0% agarose gel). The DNA was then blotted onto nitrocellulose or Gene Screen Plus (New England Nuclear) and hybridized with the ^{32}P-labeled probe under standard conditions (21).

In situ hybridization to polytene chromosomes.

Salivary glands from third instar larvae were dissected and squashed (24). To reduce the hybridization background, slides were incubated in 2X SSC

(0.3 M NaCl, 0.03 M sodium citrate) for 30 min at 58 °C, sequentially washed with 75% and 95% ethanol, dried and then acetylated by dipping for 30 min in a freshly made mixture of 250 ml of 0.1 N triethanolamine-HCl, pH 8.0 and 0.65 ml of acetic anhydride. The slides were sequentially washed with 2X SSC and then with 75% and 95% ethanol. Before hybridization the chromosomes were denatured in 0.07 N NaOH for 2 min and rinsed with ethanol. Each squash was covered with a cover slip and 2 to 3 ul of hybridization solution were introduced by capillary action. Hybridization solution contained 0.01 M PIPES, pH 6.8, 0.5 M NaCl, 0.01% w/v yeast RNA, 0.01% denatured herring DNA, 0.5 mM each of four deoxynucleotide triphosphates, 50% formamide and 4 X 10^4 cpm of DNA nick translated with ^{3}H-TTP to a specific activity of 5 X 10^6 cpm/ug . The edge of the coverslips were sealed with rubber cement and the slides were incubated at 25 °C for 24 h. After hybridization the slides were first washed for 3 X 30 min at 22 °C with 0.01 M PIPES, pH 6.8 / 0.5 M NaCl / 50% formamide, then with 75% and 95% ethanol and air-dried. The slides were dipped in Kodak NTB-2 emulsion diluted 1:1 with water and exposed for 4-7 days at 4 °C. After developing, the slides were stained with Giemsa (24).

RESULTS

Isolation, identification and sequencing of the rpA1 gene.

A cloned cDNA (hereafter refered to as "original cDNA") coding for a translationally regulated *Drosophila* mRNA had been previously isolated in our laboratory (5) and subsequently shown to code for the acidic r-protein rpA1 (3). To analyze the structure of the rpA1 gene we screened a library prepared from DNA of the Canton S strain (14) and isolated a phage containing a 20 kb insert that hybridized with the original cDNA clone. To determine which portion of the cloned DNA is transcribed into embryonic RNA, Bam HI restriction fragments covering the entire 20 kb insert were used separately as probes in hybridization to Northern blots prepared after electrophoresis of 10 ug of embryonic poly(A)+ RNA. A 2.4 kb fragment located in the middle of the insert was the only fragment to give a positive hybridization signal (data not shown). It hybridized to a 0.6 kb RNA, the same size RNA as detected by hybridization with the original cDNA. This 2.4 kb Bam HI fragment was subcloned into pBR322 and named p5D. The identity of this cloned DNA as coding for rpA1 was verified by hybrid-selected translation, followed by two-dimensional gel electrophoresis (Fig. 1). The radioactive translation product coded by the selected mRNA comigrated exactly with the stained marker rpA1 protein (see also ref. 3). As mentioned below, this turned out to be an

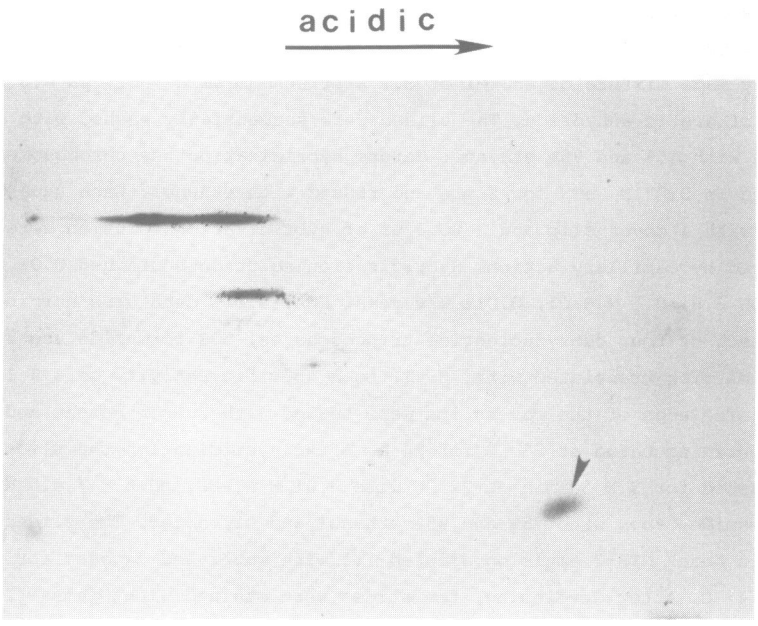

Fig. 1. DNA from the genomic clone p5D hybrid-selects a mRNA coding for rpA1. The [35S]-labeled translation products of the hybrid selected mRNA were mixed with marker r-proteins and the mixture was fractionated on a two-dimensional gel. The gel was then stained and fluorographed. The fluorogram is shown. The arrowhead points to the translation product which comigrates with the stained rpA1 marker. The spots in the upper left region of the gel are due to endogenous products of the cell-free translation system. The high molecular weight, acidic end of the gel is on the upper right corner.

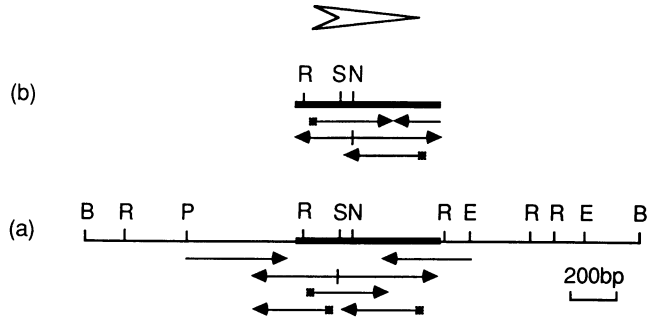

Fig. 2. Restriction map and sequencing strategy for genomic and cDNA clones of the rpA1 gene. A partial restriction map of a 2.4 kb Bam HI fragment containing the transcribed rpA1 sequence (bold portion of the line) is shown in **(a)**. The map of a full-length cDNA clone is shown in **(b)**. Regions sequenced are indicated by the arrows below the map. Arrows with stars indicate sequencing reactions primed with a synthetic oligonucleotide. The open arrowhead at the top shows the direction of transcription. B - Bam HI; E - EcoR I; N - Nru I; P - Pvu II; R - Rsa I; S - Sal I.

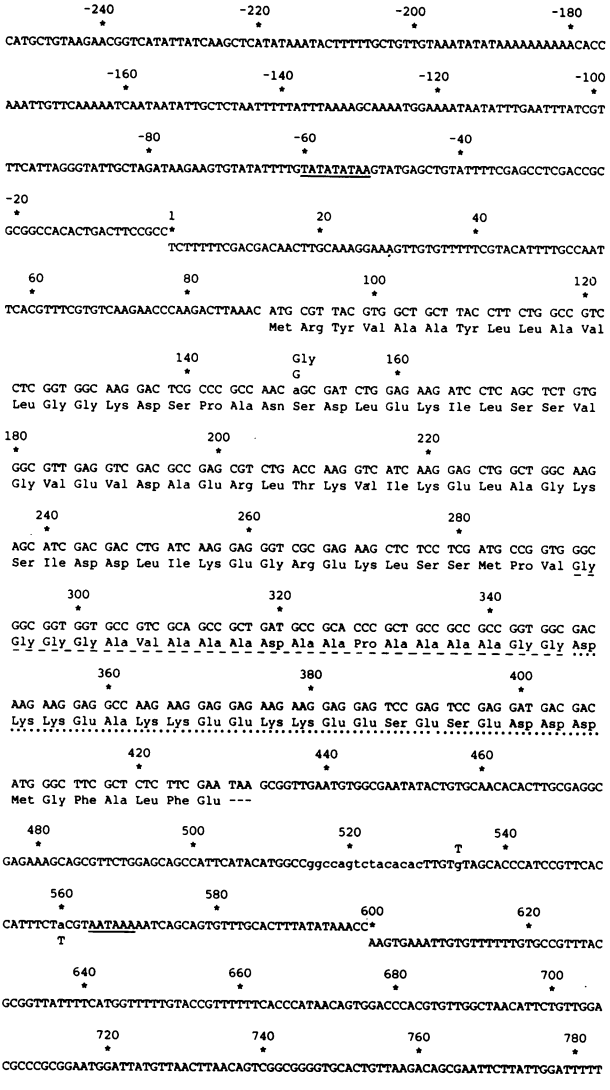

Fig. 3. Nucleotide sequence of the rpA1 gene and its deduced amino acid sequence. The nucleotide sequence of a genomic rpA1 clone from the Canton S strain was determined using the strategy outlined in Fig. 2. The nucleotide sequence of full-length cDNA clones from the Oregon R strain was also determined. Differences between the genomic and cDNA sequences are indicated by lower case letters: the cDNA has a G at position +150 causing a neutral change from Ser to Gly, a T at positions +534, +560 and misses 15 nucleotides around position +520. The beginning and the end of the transcribed sequences are indicated by a line shift. The TATA box and the polyadenylation signal sequence are underlined. A hydrophobic domain containing mostly Gly and Ala (dashed line) and a hydrophilic domain extremely rich in charged residues (dotted line) are also underlined.

TABLE I. Amino Acid Composition of Drosophila rpA1 Protein

Amino Acid	number of residues deduced from the nucleotide sequence	Content % * A	B
Asp	10	8.8	9.8
Asn	1	0.9	
Thr	1	0.9	1.4
Ser	9	7.9	7.4
Glu	14	12.4	11.5
Gln	0	0.0	
Pro	3	2.7	3.3
Gly	12	10.6	10.7
Ala	18	15.9	16.0
Cys	0	0.0	0.0
Val	8	7.1	7.8
Met	3	2.7	2.2
Leu	10	8.8	8.6
Ile	4	3.6	2.4
Tyr	2	1.8	1.9
Phe	2	1.8	2.0
His	0	0.0	0.2
Lys	13	11.5	11.0
Arg	3	2.7	3.4
Trp	0	0.0	n.d.

* A: Composition deduced from the nucleotide sequence.
B: Composition analysis of the protein isolated from embryos.
n.d.: not determined.

important control experiment since we later discovered that the original cDNA clone contains sequences unrelated to rpA1.

A restriction map of p5D was established and the coding region of the gene was sequenced (Fig. 2a and Fig. 3). The insert contains a single open reading frame spaning bases 90 to 428. The amino acid composition predicted from this reading frame matches very well with the composition of the purified protein (Table 1). This result further confirmed the identity of the cloned rpA1 gene and indicated that we had accurately determined the coding sequence of the gene. The transcription initiation site (nucleotide 1 in Fig. 3) was determined from S1 nuclease protection and primer extension experiments (Fig.4). The 3'-end of the mRNA was determined from the sequence analysis of the original cDNA and of several independently isolated cDNA clones (see below). The rpA1 gene has a TATA box at about position -60 rather than the usual position -20 to -30. A canonical AATAAA polyadenylation signal is present 40 nucleotides upstream from the mRNA end.

Sequence of rpA1 cDNA and analysis of gene structure.

The sequence of p5D predicts a primary transcript of 599 nucleotides.

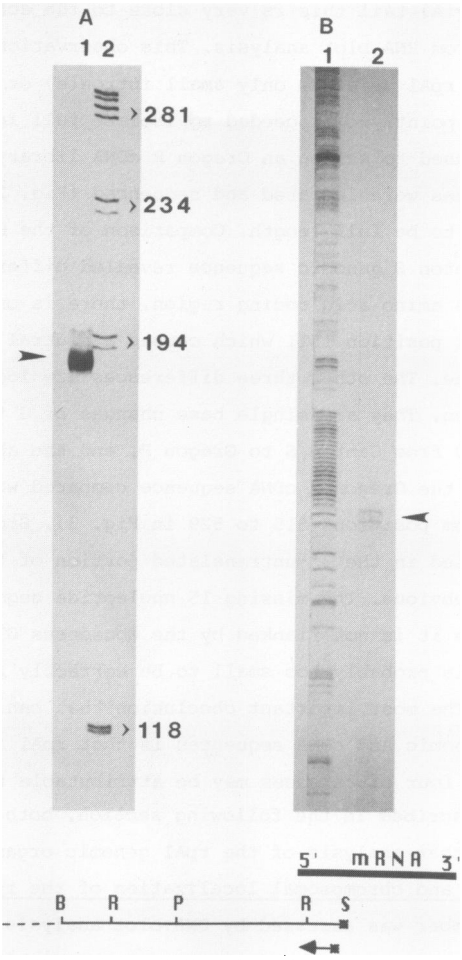

Fig.4. Determination of the transcription initiation site.
(A) S1 nuclease mapping: the DNA probe used and its relationship to the rpA1 mRNA is illustrated below the autoradiograph. The probe contained the 5'-portion of the rpA1 gene (Bam HI-Sal I fragment, Fig.2) and pBR322 sequence indicated (not to scale) by the dashed line. The probe was end labeled with [32P] (*). Compared with the size marker Hae III digested ϕX174 DNA in lane 2, the S1-resistant fragment (arrowhead) is about 190 bp in length. **(B)** Primer extension: in the drawing, the primer extension product is shown by an arrow. Reverse transcription was primed by an end-labeled (*) synthetic oligomer complementary to the sequence 46 bp upstream of the Sal I site. The size marker in lane 1 is a dideoxyguanosine sequencing reaction of a cDNA clone primed by the same synthetic oligomer. The poly(G) track on the ladder is derived from the G tail of the cDNA clone and indicates the 5' end of the cDNA. The primer extension product comigrates with the begining of the G track (arrowhead), indicating that the cDNA is of full length. The 5' end of the cDNA is 147 bp upstream from the primer which, plus 46 bp, is 193 bp from Sal I site, in good agreement with the S1 mapping data (panel A).

Allowing for the poly(A) tail this is very close to the actual size of the mRNA, as estimated from RNA blot analysis. This observation raised the possibility that the rpA1 gene has only small intron(s) or none at all. To further examine this point, we proceeded to isolate full length rpA1 cDNA clones. The p5D was used to screen an Oregon R cDNA library (15). Three independent cDNA clones were isolated and sequenced (Fig. 2a. and Fig. 3). All three were found to be full length. Comparison of the Oregon R cDNA sequence with the Canton S genomic sequence revealed differences at four positions. Within the amino acid coding region, there is only one single base change from A to G at position 151, which causes a neutral amino acid change from serine to glycine. The other three differences are located at the 3'-untranslated region. They are single base changes of G to T and A to T at positions 534 and 560 from Canton S to Oregon R, and the absence of a stretch of 15 nucleotides in the Oregon R cDNA sequence compared with the Canton S genomic sequence (from positions 515 to 529 in Fig. 3). Since these differences are located in the 3'-untranslated portion of the mRNA, their significance is not obvious. The missing 15 nucleotide segment is probably not an intron because it is not flanked by the consensus GT/AG nucleotides (25) and because it is probably too small to be correctly removed by lariat formation (26, 27). The most important conclusion that can be drawn from the comparison of the genomic and cDNA sequences is that rpA1 is most likely an intronless gene. The four differences may be attributable to strain polymorphisms. As described in the following section, both wild type strains were compared in further analysis of the rpA1 genomic organization.

Genomic organization and chromosomal localization of the rpA1 gene.

The gene copy number was assessed by DNA blot analysis of fly DNA and by in situ hybridization to polytene chromosomes. Southern blot analysis using a full length cDNA probe, revealed significant differences between Canton S and Oregon R strains. While mostly single bands were obtained with Oregon R DNA (Fig. 5A), Canton S DNA yielded double bands when cut with several restriction enzymes (Fig. 5B). At this level then, the rpA1 gene appears to be single copy in the Oregon R genome. The Canton S pattern may be attributed to restriction site polymorphism either between two allelic copies of a single gene or between two non-allelic copies of duplicated genes in the genome. To distinguish between the two possibilities, we repeated the Southern analysis of Canton S genomic DNA several months later. Surprisingly, the DNA prepared this time showed only single bands when digested with either Bam HI or EcoR I (Fig. 5C, lanes 5b and 6b) while the DNA prepared months ago

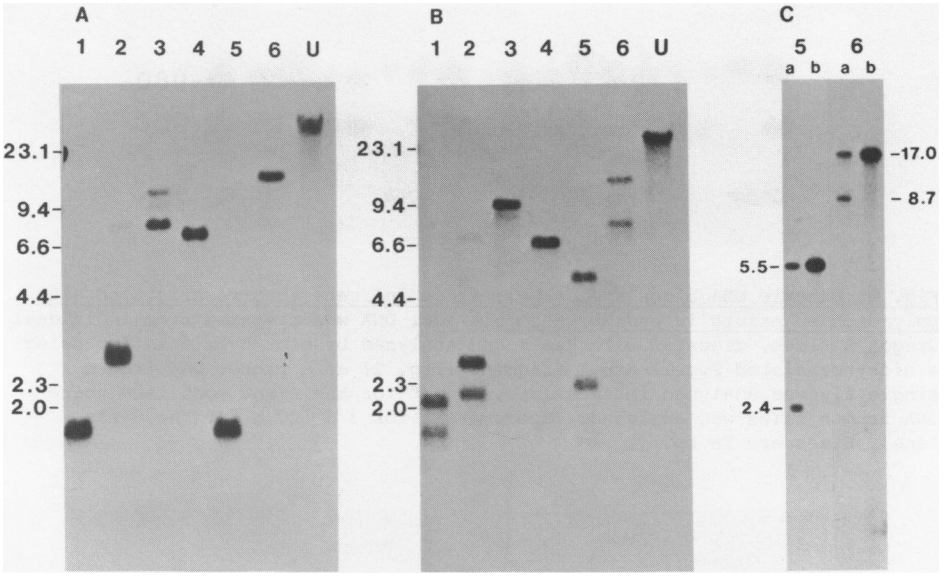

Fig. 5. Analysis of the genomic organization of rpA1 in two wild type strains of D. melanogaster. Total fly DNA (10 ug per lane) from the Oregon R **(A)** or Canton S **(B)** strains was digested to completion with different restriction enzymes or enzyme combinations, fractionated on a 0.8% agarose gel, transfered to nitrocellulose, and hybridized with a [32P]-labeled full-length rpA1 cDNA. Digestion of the DNA in the different lanes was as follows. 1 - Eco RI/Bam HI; 2 - Eco RI/Kpn I; 3 - Hind III; 4 - Xho I; 5 - Bam HI; 6 - Eco RI; U - undigested. The size marker (given in kb) was Hind III-digested lambda DNA. Panel C (lanes 5 and 6) shows the result of a similar experiment as that shown in lanes 5 and 6 of **(B)**, performed several months later. The same DNA as in **(B)** 5 and 6 was used in **(C)** 5a and 6a. The DNA used in **(C)** 5b and 6b was prepared several months later. The actual sizes of the bands are marked in **(C)**; they are of the same size as the bands in **(B)** 5 and 6. The lower 2.4 kb Bam HI and 8.7 kb EcoR I bands correspond to the sizes predicted by the restriction map of the cloned rpA1 genomic sequence (not shown).

again revealed the double band pattern (Fig. 5C, lanes 5a and 6a). The results in Fig. 5C suggest that as in the Oregon R strain, the rpA1 gene is in fact single copy in the Canton S genome. The double band pattern observed with the earlier DNA preparation may be due to allelic polymorphism. Since our Canton S stock was kept as a small number of flies between the two experiments, a biased propagation may have led to the loss of one of the allelic forms. Additional evidence shows that allelic polymorphism exists in our inbred stock: when DNA from single Oregon R flies was digested with the enzyme Rsa I and analyzed by Southern hybridization with a rpA1 probe, three different restriction patterns were observed among individual flies,

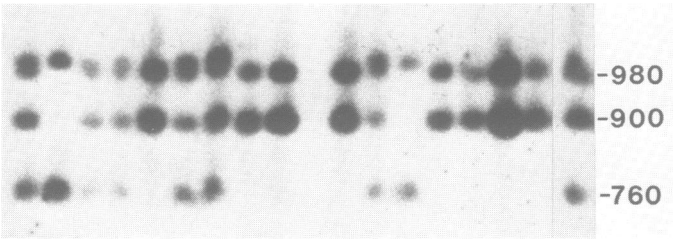

Fig. 6. Genomic DNA blots from single flies indicate that restricition site polymorphism exists in an inbred population. DNA was prepared from individual Oregon R flies, digested with Rsa I and analyzed by blot hybridization using a nick-translated Pvu II-Nru I fragment (Fig. 2) as a probe. DNA from a single fly was analyzed in each lane, except for the right most lane where DNA from 6 flies was analyzed. Exposure was for 3 d (20 h for the 6-fly lane). Sizes are in bp.

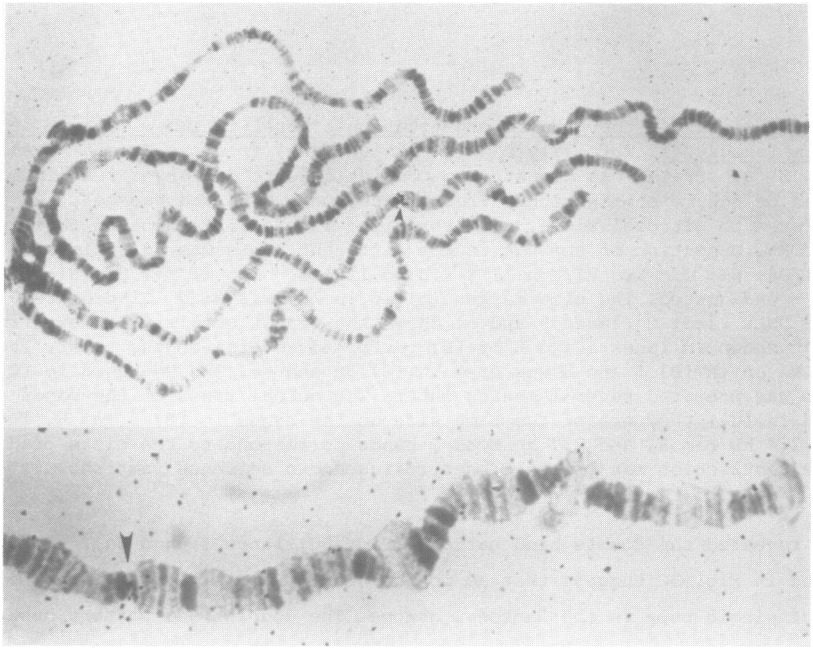

Fig. 7. In situ hybridization of an rpA1 probe to polytene chromosomes. Salivary gland polytene chromosome squashes from Canton S flies were hybridized with [3H]-labeled p5D plasmid containing the rpA1 genomic sequence. The autoradiograms produced silver grains (arrowhead) over a single band at 53CD on the right arm of the second chromosome. The bottom panel depicts under higher magnification the precise location of the silver grains. The same result was obtained with chromosomes from the Oregon R strain (not shown).

```
                          20                  40
Fly RPA1        MRYVAAYLLAVLGGKDSPANSDLEKILSSVGVEVDAER-LTKVIKELAGK-SIDDLIK
Rat P2          MRYVAsYLLAaLGGnsnPsakDiaKILdSVGiEaDdERkLnKVIsELnGK-nIeDvIa
Yeast YPA1      MkYlAAYLLlVqGGnaaPsaaDikavveSVGaEVDeaR-inellssLeGKgSleeiIa
Shrimp eL12     MRYVAAYLLAaLsGnadPstaDiEKILSSVGiEcnpsq-LqKVmnELkGK-dleaLIa
Shrimp eL12'    ------------------------------------------------------
                ::::::::::.:::...:...:..:::.:::.:.:..:-:.::..::.::-.....:-

                 60                  80                  100
Fly RPA1        EGREKLSSMPVGGGGAVAAADAAPAAAAGGDKKEAKKEEKKEESESEDDDMGFALFE
Rat P2          qGvgKLaSvPaGGavAVsAApgAaApAAGsa---paaaEeKEESEekkDeMGFgLFd
Yeast YPA1      EGqkKfatvPtGGassagpAsAg-AAAgGGDaaEeKeEEaKEE---sDDDMGFgLFd
Shrimp eL12     EGqtKLaSMP--tGGApAAAaggaAAApaaeaKEAKKEEKKEESEeEDeDMGFgLFd
Shrimp eL12'    -----------sGvGAapAAggAaAAteapaaKEeKKEEKKEESEeEDeDMGFgLFd
                ::..::.:.:.::.::..::..:.:::.:..-::.:::::::::.:::::::.::.
```

Fig. 8. Homology between Drosophila ribosomal protein rpA1 and other eucaryotic "A type" r-proteins. The Drosophila (fly) rpA1 amino acid sequence was aligned with the rat P2 (8), yeast YPA1 (9), A. salina (shrimp) eL12 and eL12' (10, 27) (eL12 and eL12' are coded by different genes) r-protein sequences. For eL12', only the C-terminal segment which shares homology with other "A" proteins is shown. The amino acids that differ from the fly rpA1 sequence are given with lower case letters. A few gaps (indicated by hyphens) have been introduced into the sequences to maximize the homology. A summary of the analysis is given in the lower line and includes the following categories: (:) perfect match; (.) conservative change; (-) no match. When comparisons were made with the fly sequence, assignment to a given category was made when two out of three (or four when including the shrimp eL12' sequence) satisfied the above criteria.

indicating that allelic polymorphism within our inbred population indeed exists (Fig. 6). In situ hybridization to polytene chromosomes gave a single hybridization band at 53CD in flies of both strains (Fig. 7), consistent with the single gene interpretation. The location of the hybridization band at 53CD was unexpected because previous experiments using the original cDNA probe had indicated that rpA1 is located at 39CD (5). This apparent discrepancy was clarified when the original cDNA insert was sequenced. In addition to the rpA1 sequence, the original cDNA clone contained a stretch of DNA of unknown origin (data not shown), possibly a cloning artifact. Thus, it is possible that under the hybridization conditions used previously, the stretch of foreign DNA led to the hybridization at 39CD. It is not clear why the original cDNA did not give a signal at 53CD in the previous experiments.

Acidic rpA1 is an "A" type eucaryotic ribosomal protein.

The rpA1 gene codes for a protein of unusual composition and sequence (Table 1 and Fig. 3). About 16% of its residues are alanines. Of the 19 amino acid stretch coded by nucleotides 291 to 347, 10 are alanines and 6 are glycines. This rather hydrophobic stretch is immediately followed by an

extremely hydrophilic portion of the protein where 17 out of 20 residues are charged. The rpA1 protein has no histidine, cysteine or tryptophan. As expected from its migration on two-dimensional gels (Fig. 1), this is one of the few r-proteins that are acidic, having a calculated net negative charge of -8. A search of the Protein Resource Indentification Data Base revealed that the amino acid sequence of Drosophila rpA1 has significant homology with all four members of the "A" family of eucaryotic acidic r-proteins for which amino acid sequence data is available (8, 9, 10, 28). Drosophila rpA1 has an overall identity in 61.1%, 46.5%, 55.3% of its amino acids with A. salina eL12, S. cerevisae YPA1, and rat P2 r-proteins respectively and 60.9% identity with the 46 C-terminal residues of A. salina eL12' (Fig. 8) (eL12 and eL12' are coded by different genes). If conservative amino acid changes are included, the homology in each case is better than 90%. Moreover, "A" type proteins are characterized by a net negative charge (most r-proteins are positively charged) and a peculiar amino acid composition: high alanine content (about 20%), few aromatic amino acids, 2 or 3 arginines, and no cysteine nor tryptophan. Drosophila rpA1 fits these characteristics very well (Table 1). At the nucleotide sequence level, rpA1 shares 64.4% homology with A. salina eL12 (29). No significant homology was found between Drosophila rpA1 and other r-proteins.

DISCUSSION

rpA1 (this report) and rp49 (30) are the only Drosophila r-protein genes whose nucleotide sequences have been determined. A list of other eucaryotic r-protein genes that have been cloned can be found in ref. 31. Among the noteworthy features of the rpA1 gene are the facts that the gene is intronless, that it codes for a protein that belongs to a class of conserved r-proteins whose function is partially known, and that its expression is selectively regulated during embryonic development of Drosophila.

Conservation of the rpA1 sequence between two Drosophila strains is indicated by the nearly identical coding sequences of the Canton S genomic clone and Oregon R cDNA clone, while restriction site polymorphism has been observed outside of the gene. Genetic polymorphism between two Xenopus strains has been reported for r-protein L1 (32). Surprisingly, polymorphism flanking the rpA1 gene was also observed within each inbred Drosophila strain after digestion with a number of restriction enzymes (Figs. 5 and 6). The Xenopus L14 gene also appears to be polymorphic among individual frogs of the same strain (33).

The other Drosophila r-protein gene that was sequenced (30) does contain an intron, and so do most Drosophila and other eucaryotic genes. Furthermore, most yeast r-protein genes contain an intron, despite the fact that the majority of its genes are intronless. The significance of finding a Drosophila intronless r-protein gene is not evident at the moment. The presence and absence of an intron in different r-protein genes suggest that pre-mRNA processing could not be the only point at which r-protein synthesis is coordinated. Recent studies with yeast and Xenopus have shown that the expression of r-protein genes can be regulated at at least three levels: pre-mRNA processing, translation and protein turnover (2, 34).

The presence of a TATA box 60 bp upstream from the rpA1 transcription initiation site is unusual. Many r-protein and other "housekeeping" genes do not have a well defined TATA element (35, 36) while most eucaryotic genes have such element at position -30. Interestingly, Drosophila rp49 also has a TATA box at about position -50, farther upstream from the "normal" position. The seven nucleotides preceding the AUG translational initiation signal of rpA1 - CTTAAAC - match quite well the consensus sequence - ANNC(A)AA(C)A(C) - that has been derived from 23 sequenced Drosophila genes (N represents any nucleotide; values in parenthesis represent a "co-consensus" with the preceding nucleotide) (Doug Cavener, personal communication).

Comparison of the deduced rpA1 amino acid sequence with the sequence of other r-proteins, identified rpA1 as a member of the eucaryotic "A" family r-proteins. "A" proteins are extremely conserved in evolution, having representatives in an archebacterium (H. cutirubrum), in a fungus (yeast), in an invertebrate (A. salina), in a mammal (rat) and now in Drosophila. The "A" group of r-proteins has a counterpart in a class of procaryotic acidic r-proteins of which E. coli r-protein L7/L12 is a prototype (7) (L7 differs from L12 only in the acetylation of the N-terminal serine). In E. coli, L7/L12 participates in the binding of initiation, elongation and termination factors of protein synthesis (7). In view of the conserved nature of this class of proteins in evolution, it is reasonable to assume that "A" proteins have similar functions. In fact, A. Salina eL12 has already been shown to be essential in the GTP driven elongation process (13). Finally, knowledge of the detailed structure of rpA1 provides us with an essential tool with which to investigate the molecular basis for the specific regulation of r-protein mRNA translation during Drosophila embryonic development.

ACKNOWLEDGEMENTS

We thank Dr. T. Rosenbery for help with the amino acid composition analysis and Drs. J. R. Thompson, P. J. Harte and K. Burtis for help with the computer analysis. This work was supported by grants from the National Institutes of Health and the National Science Fundation.

*To whom correspondence should be addressed

[1]Present address: Department of Biochemistry, Tianjing Medical College, Tianjing, People's Republic of China

REFERENCES

1. Kim, C.H. and Warner, J.R. (1983) J. Mol. Biol. 165, 79-89.
2. Warner, J.R., Mitra, G., Schwindinger, W.F., Studeny, M. and Fried, H.M. (1985) Mol. Cell. Biol. 5, 1512-1521.
3. Kay, M.A. and Jacobs-Lorena, M. (1985) Mol. Cell. Biol. 5, 3583-3592.
4. Nomura, M., Gourse, R. and Baughman, G. (1984) Ann. Rev. Biochem. 53, 75-117.
5. Fruscoloni, P., Al-Atia, G.R. and Jacobs-Lorena, M. (1983) Proc. Natl. Acad. Sci. USA 80, 3359-3363.
6. Al-Atia, G.R., Fruscoloni, P. and Jacobs-Lorena, M. (1985) Biochemistry 24, 5798-5803.
7. Wool, I.G. (1979) Ann. Rev. Biochem. 48, 719-754.
8. Lin, A., Wittmann-Liebold, B., McNally, J. and Wool, I.G. (1982) J. Biol. Chem. 257, 9189-9197.
9. Itoh, T. (1981) Biochim. Biophys. Acta 671, 16-24.
10. Amons, R., Pluijms, W. and Moller, W. (1979) FEBS lett. 104, 85-89.
11. Moller, W. (1974) In Nomura, M., Tissieres, A. and Lengyel, P. (eds), Ribosomes, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 711-731.
12. Moller, W., Schrier, P.I., Maassen, J.A., Zantema, A., Schop, E. and Reinalda, H. (1983) J. Mol. Biol. 163, 553-573.
13. van Agthoven, A.I., Maassen, J.A. and Moller, W. (1977) Biochem. Biophys. Res. Commun. 77, 989-998.
14. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C. and Quon, D. (1978) Cell 15, 687-701.
15. Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) Cell 40, 37-43.
16. Ruddell, A. and Jacobs-Lorena, M. (1984) Dev. Biol. 105, 115-120.
17. Ricciardi, R.P., Miller, J.S. and Roberts, B.E. (1979) Proc. Natl. Acad. Sci. USA 76, 4927-4931.
18. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
19. Dente, L., Cesareni, G. and Cortese, R. (1983) Nucl. Acids Res. 11, 1645-1655.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
21. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
22. Domdey, H., Apostol, B., Lin, Ren-Jang., Newman, A., Brody, E. and Abelson, J. (1984) Cell 9, 611-621.
23. Hunkapiller, M.W., Lujan, E., Ostrander, F. and Hood, L.E. (1983) Methods in Enzymol. 91, 227-236.

24. Pardue, M.L. and Gall, J.G. (1975) Methods in Cell Biology 10, 1-16.
25. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
26. Keller, W. (1984) Cell 39, 423-425.
27. Upholt, W.B. and Sandell, L.J. (1986) Proc. Natl. Acad. Sci. USA 83, 2325-2329.
28. Amons, R., Pluijms, W., Kriek, J. and Moller, W. (1982) FEBS lett. 146, 143-147.
29. Maassen, A., Schop, E.N., Brands, Jolanda H.G.M., Van Hemert, F.J., Lenstra, J.A. and Moller, W. (1985) Eur. J. Biochem. 149, 609-616.
30. O'Connell, P. and Rosbash, M. (1984) Nucl. Acids Res. 12, 5495-5513.
31. Jacobs-Lorena, M. and Fried, H. (1987) In Ilan, J. (ed), Translational Regulation of Gene Expression, Plenum Press, New York, in press.
32. Bozzoni, I., Tognoni, A., Pierandrei-Amaldi, P., Beccari, E., Buongiorno-Nardelli, M. and Amaldi, F. (1982) J. Mol. Biol. 161,353-371.
33. Bozzoni, I., Beccari, E., Luo, Z.X. and Amaldi, F. (1981) Nucl. Acids Res. 9, 1069-1086.
34. Bozzoni, I., Fragapane, P., Annesi, F., Pierandrei-Amaldi, P., Amaldi, F. and Beccari, E. (1984) J. Mol. Biol. 180, 987-1005.
35. Teem, J.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., van Raamsdonk-Duin, M.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D., Fried, H. and Rosbash, M. (1984) Nucl. Acids Res. 12, 8295-8312.
36. Dynan, W.S. (1986) Trends in Genetics 2, 196-197.