***Modulo*, a new maternally expressed *Drosophila* gene encodes a DNA-binding protein with distinct acidic and basic regions**

Eric Krejci, Veronique Garzino, Charles Mary, Naima Bennani and Jacques Pradel*

Laboratoire de Génétique et de Biologie, Cellulaires du Centre National de la Recherche Scientifique, Luminy, Case 907, 13288 Marseille cedex 09, France

Received July 26, 1989; Revised and Accepted September 15, 1989

EMBL accession no. X15702

ABSTRACT

We have cloned, following an immunological screen of an expression library, five cDNA clones encoding the *modulo* antigen, a DNA-binding protein differentially expressed during *Drosophila* development. In addition a series of overlapping cDNA and genomic clones were also isolated. This protein is the product of a 2.2 kb mRNA that is encoded by a single genetic locus (100F). Analysis of the complete 544 amino-acid sequence, deduced from nucleotide sequence of cDNAs, shows that the polypeptide exhibits a primary structure with distinct charged regions, a modular structure found in several eukaryotic nuclear proteins, either transcription regulators or structural factors. The amino and carboxyl termini are rich in basic residues. The first third of the sequence contains a long domain comprised almost entirely of glutamic and aspartic acid residues. A typical cAMP dependent phosphorylation site and five potential glycosylation sites have been detected in the amino-acid sequence. Computer searches fail to reveal any significant homology with known proteins. Developmental pattern of transcription of the *modulo* gene indicates that messengers are maternally provided to the embryos and that zygotic transcription is required during subsequent development.

INTRODUCTION

Differential gene expression is controlled in part at the level of transcription following interaction of nuclear proteic factors and the control regions of genes (1, 2). Recent work, based on gene sequence analysis and on construction of chimeric gene products (3, 4), has provided new information on the nature of eukaryotic transcriptional activators. These proteins possess distinct domains involved in at least two different functions ; a basic DNA-binding domain, with a typical structural motif like the homeodomain (5), zinc finger (6) or amphipatic helix-loop-amphipatic helix (7), positions the protein on DNA so that the other, an activating domain which is often acidic, can interact with other protein(s) to promote initiation of transcription (3, 8).

Several nuclear proteins have been identified in *Drosophila melanogaster*, that probably regulate genes or gene sets during development. For instance, the products of segmentation and homeotic genes, which define the pattern of metameric units during embryogenesis (9, 10), are nuclear proteins and are probably involved in gene regulatory functions (11, 12). Moreover some genes encoding DNA-binding proteins have been described, that exhibit specific developmental transcription pattern (13) or that interact with other genes during embryonic development (14, 15). Several of these factors, such as the engrailed (16), Sry (13) or su(Hw) (15) gene products, present the modular structure of transcriptional activators, i. e. an acidic domain and a DNA-binding motif.

Various approaches have been developed to isolate new genes involved in transcription regulation during *Drosophila* embryogenesis. DNA sequences from previously indentified

genes have been used to obtain additional homeobox containing genes (17) or zinc finger protein coding genes (18). Detection of genomic regulatory elements by random insertion of reporter genes has also led to isolation of new genes that play a crucial role in development (19, 20). Another approach consists of looking directly for DNA-binding proteins, displaying a specific temporal and spatial expression suggestive of gene regulatory function. We have recently reported (21) a collection of monoclonal antibodies against nuclear proteins differentially expressed during *Drosophila* development.

The antigen recognized by monoclonal antibody LA9 (Mab LA9), first detected at the cellular blastoderm stage in all the nuclei of the cellular epithelium, progressively restricts its localization during embryogenesis and by the end of organogenesis is localized in the nuclei of the gut and associated structures (21). The antigen, the protein *modulo* (*mod*), is a chromosomal component that binds DNA and also interacts with other nuclear proteins (Garzino et al., unpublished). We report here the isolation of cDNA and genomic clones encoding this antigen and the determination of the complete amino-acid sequence deduced from nucleotide sequencing. This new gene, the gene *modulo* (*mod*), maps at 100 F, the telomeric region of chromosome 3R. *Mod* gene product is an acidic protein that exhibits a modular structure consisting of three distinct regions, two basic regions at the carboxyl and amino termini and a strongly acidic region in the first third of the sequence.

MATERIALS AND METHODS

DNA preparation, digestion by restriction enzymes, electrophoresis and nucleic acid blotting were carried out by standard procedures (22). Labelled nucleotides ^{32}P XTP (3000 Ci/mMole), ^{32}P ATP (5000 Ci/mMole) and ^{35}S thio dATP (1000 Ci/mMole) were obtained from Amersham.

Nuclei purification from *Drosophila* embryos and immunofluorescence assays on cryostat tissue section have been performed as previously described (21).

Libraries : The lambda gt 11 cDNA expression library used here was a gift of Dr. M.P. Lee (Duke University). This library was prepared from total poly (A⁺) RNA isolated from *D. melanogaster* embryos (23). The lambda EMBL4 genomic DNA library (*Drosophila* Oregon R) was a gift of H. Lehrach.

Isolation and characterization of cDNA and genomic clones : The lambda gt 11 cDNA library was screened by using the Mab LA9 purified from ascites fluid. Phages were plated on *E. coli* Y1090 at a density of 100,000 phages per 150 mm plate. Following incubation at 42°C during 3 h 30, each plate was overlaid with an IPTG (Isopropyl- β -D-thiogalactopyranoside) impregnated nitrocellulose filter and incubated at 38°C during 3 h to obtain one replica or 2 $\times$ 2 h when two replicas were needed (24). Nitrocellulose filters were incubated for 1 h at 37°C in 5 % Blotto (PBS, 0.1 % Triton X100, 5 % dry defatted milk) and then placed overnight at 4°C with Mab LA9 in 0.5 % Blotto. Positive plaques were detected with an alkaline phosphatase-conjugated second antibody (Promega).

The genomic library was screened by with the Eco RI fragment (474–1816) purified from a positive lambda gt 11 cDNA clone (LA913).

The lambda gt 11 cDNA library was rescreened with two cDNA probes (fragments : EcoRI-EcoRI (474–1816) and EcoRI-EcoRI (–81–474) from phage LA913 and with one genomic probe (fragment : BamHI-MluI (–70), approximatively 1 kb).

Production of fusion proteins and of polyclonal antisera : The lysogens in *E. coli* Y 1089 were easily selected with a phage cInt which cannot be inserted in the chromosome (a gift of Dr. M. Murgier, Marseille). The fusion proteins were produced as described by

Figure 1: Expression and serological identification of fusion proteins in lysogenized *E. coli*. Total proteins from *E. coli* Y1089 lysogenized with the five recombinant phages obtained from immunological screening of cDNA expression library (lane 1 to 5 in panels A and B, lane 1 in panel C) and nuclear proteins isolated from *Drosophila* embryos (lane 2 in panel C) were run on SDS/5%PAGE (5 μ g per lane) and blotted onto nitrocellulose. Blots were probed with anti- β -galactosidase (panel A) and with Mab LA9 (panels B and C). Lane C1 is an overloaded lane (20 μ g) of proteins from *E. coli* lysogenized with LA 913 phage.

Huynh *et al.* (24). Briefly, the lysogens were grown in LB medium supplemented with 10 μ g of ampicillin per ml at 32°C until cell density reached 5×10^6 . The cells were induced by shifting the temperature to 42°C and bringing the medium to 1 mM IPTG. After an induction period of 15 min., the cells were incubated at 38°C for another 45 min and then harvested by centrifugation. The bacteria were broken by sonication and proteins were separated by SDS-PAGE. The immunological activity of these proteins was analysed on Western blot after transfer to nitrocellulose membrane.

Fusion proteins were purified from lysogen lysates by immunoaffinity chromatography (Mab LA9 coupled ACA 22 ultrogel, IBF) and SDS-PAGE. Balb/c mice were immunized by repeated injections of 0.1 mg of gel purified fusion protein blotted onto nitrocellulose: three injections each two weeks and a last boost one week before intracardiac blood taking off. Antibodies directed against β -gal epitopes were removed by chromatography on β -gal columns (Promega).

DNA cellulose chromatography. 80 μ g of protein extracted from embryo (0–18 h) nuclei, and approximately 1 μ g of immunopurified β gal-*mod* fusion protein were loaded onto 0.4 ml columns containing single-stranded or double-stranded DNA cellulose (Pharmacia) (25) equilibrated in buffer A (15 mM Tris-HCl (pH 8.0), 1 mM DTT, 0.05 % NP₄₀, 5 mM EDTA and a cocktail of protease inhibitors containing : 0.5 mM PMSF, 3 μ g ml⁻¹ aprotinin and 2 μ g ml⁻¹ antipain, leupeptin and pepstatin). Fusion protein was immunopurified as above mentioned ; however after SDS-PAGE, the gel was sliced and the fusion protein was electrically eluted and dialysed against 10 mM Tris-HCl (pH 8.0), 1 mM PMSF. The columns were washed with buffer A and the proteins eluted with buffer A containing 50 mM, 100 mM, 500 mM and 1 M NaCl. Proteins were precipitated in

Figure 2: Immunodetection of *Drosophila mod* antigen by using an antiserum directed against the fusion protein produced by LA913 phage. A : Western blot analysis. *Drosophila* nuclear proteins (lane 1, 3) and fusion protein (lane 2, 4) were run on SDS/5%PAGE, blotted onto nitrocellulose and probed with Mab LA9 (1, 2) or with the polyclonal mouse antiserum (3, 4). The antiserum recognized in lane 3 a single component of 78 kD that co-migrates with *Drosophila mod* antigen. Molecular weight standards are indicated on the left. B: Immunofluorescence analysis. Frozen embryos (12 h) sections were stained with the antiserum (panel 1) and with Hoechst 33342 dye to visualize nuclei (panel 2). Immunostaining of nuclei of the gut and associated structures is reminiscent of that observed with Mab LA9 (21).

5 % trichloroacetic acid, separated on SDS-PAGE, blotted onto nitrocellulose and detected on the western blot by reaction with Mab LA9.

RNA purification and Northern blots: *Drosophila* Oregon R wild type eggs were collected and homogenized in thiocyanate and the RNA were banded in a CsCl gradient according to Fyrberg *et al.* (26). Poly (A⁺) RNA were purified by oligodT-cellulose chromatography and fractionned on formaldehyde 1.2 % agarose. RNA was transferred to nitrocellulose by blotting and then probed with LA913 Eco RI (474–1816) fragment labelled by random priming of DNA (27). The filters were washed to a stringency of $0.1 \times \text{SSC}$ at 55°C.

Total RNA isolated from isolated ovaries, staged embryos (0–1h 30, 2h–3h, 4h–6h, 12h–18h) larvae (third instar), pupae and virgin females were dot-blotted onto nitrocellulose (10 µg each dot) and probed with ³²P labelled LA 913 1,4 EcoRI fragment.

Genomic Southern blots analysis : Genomic DNA was digested with restriction enzymes, and separated on 0.8% agarose gel, transferred to nylon membranes, Hybond N

(Amersham), or to gene screenTM plus (NEN Research product) according to the procedures described by the manufacturers.

In situ localization of the gene mod: In situ localization on the polytene chromosomes was done according to Engels (28) using a biotinylated cloned genomic DNA probe (EMBO 4 LA1). A polyalkaline phosphatase enzyme-linked immunoassay (Bethesda Research Laboratories) was used to identify the hybrids.

Nucleotide sequence analysis

Purified cDNAs and genomic fragments of DNA were cloned in bluescript KS⁺ (Stratagene). The cDNA fragments of LA913, SacI (in the Lac Z sequence)-XbaI (1071) and EcoRI-EcoRI (474–1816), were cloned in bluescript KS⁺ (Stratagene) in the two orientations. The deletions were generated with exonuclease III according to Yannish-Perron *et al.* (29). The transformants were grown with M13K07 as helper to produce single strand phage (30). The culture was centrifuged and 40 μ l of the supernatant mixed with 6 μ l of SDS gel loading buffer (0.05 % bromophenol blue, 0.2 M EDTA, pH 8.0, 50 % glycerol, 1 % SDS). The deleted subclones were selected after analysis on agarose gel. The dideoxynucleotide sequencing technique of Sanger *et al.* (31) was used with the Sequenase (United States Biochemica Corp.), ³⁵S dATP and 40 cm gel (32).

The programs written by Bellon (33) were used to compile sequence data and restriction maps. Screening protein sequence data banks (Genepro and NBRF) was made by connecting to Citi 2 calcul center (Paris).

RESULTS

Isolation of cDNA clones that code for the mod antigen.

The cDNA library in lambda gt11 was prepared by Nolan *et al.* (23) from poly (A⁺) RNA isolated from *Drosophila melanogaster* embryos. Screening of 8×10^5 cDNA recombinant phages with Mab LA9 identified five plaques that on IPTG induction produce β -galactosidase fusion proteins that were antigenically related to the nuclear antigen *mod*. All five recombinant phages contain inserts between 1.7 and 2 kb. Restriction endonuclease digestion experiments and Southern blot hybridization analysis (not shown) indicated that the five inserts share the same 1.4 kb EcoRI fragment (see Fig. 5B).

Evidence that the cDNA clones code for mod protein

E. Coli Y 1089 was lysogenized with the five recombinant phages and the proteins obtained from the induced lysogenized stains were analyzed on western blot. All the recombinants reacted with an anti- β -galactosidase serum (Fig. 1a) and with Mab LA9 (Fig. 1b). Mab LA9 recognized the high molecular weight polypeptides. The molecular weight of the largest recognized polypeptide is 190 kD. This polypeptide results from the fusion of the β -galactosidase (116 kD) with a polypeptide of about 75 kD produced from the 2 kb insert in LA 913.

It is noteworthy that besides the high molecular weight fusion protein, Mab LA9 recognized another polypeptide produced in small amounts in IPTG induced lyzogen lLA913. This polypeptide of 78 kD was not recognized by anti- β -galactosidase antibodies and exactly co-migrated with *Drosophila mod* antigen (Fig. 1C).

To ascertain the validity of the cDNA clones isolated from the immunological screen, we attempted to compare the NH₂ terminal amino-acid sequence of *mod* protein to that of deduced from nucleotide sequence of the 5' part of LA913 cDNA. This could not be done because the NH₂ terminal of the protein, immunopurified from embryos, was blocked. So, we raised polyclonal antisera against fusion protein to set up that the polypeptide

Figure 3: Column chromatography of *Drosophila mod* protein and of fusion protein on single-stranded and double-stranded DNA-cellulose. Purified bacterial β gal-*mod* fusion protein (A, B), β galactosidase (C) and nuclear lysates of *Drosophila* embryos were passed through single-stranded (A, D) and double-stranded (B, C, E) DNA cellulose columns and eluted with by increasing salt concentration (lanes : (1) flow through ; (2) 0 mM NaCl wash ; (3), (4) : 1 bed volume, 50 mM NaCl ; (5), (6) 1 bed volume, 100 mM NaCl ; (7), (8) 1 bed volume, 500 mM NaCl ; (9) 1 bed volume, 1 M NaCl). Proteins in each fraction were precipitated in 5 % trichloroacetic acid, resolved in Laemmli buffer, separated on SDS-PAGE and electroblotted onto nitrocellulose. Blots were probed with Mab LA9. Molecular weight standards are indicated on the left.

produced by the cDNA inserts corresponded to *mod* antigen. Mice were injected with fusion protein purified from two lysogens (ILA913 and ILA909) and the resulting antisera, depleted of antibodies against β -galactosidase, were assayed on Western blot and immunofluorescence experiments.

Mab LA9 recognized a polypeptide of 78 kD on Western blot of nuclear proteins and stained the nuclei of endo-mesodermal cells in immunofluorescence assays on sections of late embryos (21). On Western blots, the purified antisera reacted specifically with only one nuclear protein of 78 kD, the same apparent molecular weight as the *mod* antigen (Fig. 2A). Immunofluorescence data shown in Fig. 2B indicate that in 12 h old embryos, the cognate antigen located in nuclei of gut and mesodermal cells ; by contrast, nuclei of nerve cord and ectodermal cells were not stained. This in situ distribution is the same as that observed for *mod* antigen (21).

Figure 4: RNA transfer blot hybridization of *mod* transcript. Poly (A⁺) RNA (1 µg) from *Drosophila* embryo (0–18 h) was fractionated in 1% agarose gel containing formaldehyde and transferred to nitrocellulose membrane. The filter was probed with the 1.4 kb EcoRI-EcoRI fragment of LA913 insert. The molecular weight of *mod* mRNA was determined through comparison with RNA size marker (0.24–9.5 kb RNA ladder, Bethesda Research Laboratory).

DNA binding activity of *mod* *Drosophila* protein and of bacterial β gal-*mod* fusion protein. The DNA binding activities of the protein produced by lysogen ILA913 and of *Drosophila mod* protein have been analyzed by chromatography on DNA cellulose. Nuclear lysates isolated from embryos and immunopurified fusion protein were passed through single or double stranded calf thymus DNA columns. Columns were eluted with NaCl from 0 to 1 M and each fraction then analyzed by SDS-PAGE and western blot. As shown in Fig. 3 (D, E) the *Drosophila* protein was maximally eluted from the DNA columns at a NaCl concentration of 0.5 M. About 60 % of *mod* protein reproducibly bound to either column, with the rest of the protein found in the flow-through fraction. These observations have at least two possible explanations : either the protein has variable DNA affinity, depending on post-translational modification or interaction with other factors ; or the protein binds to a limited number of DNA sites that are easily saturated. In favour of the second possibility, we observed that *mod* antigen contained in the flow through bound to DNA cellulose upon repassage (not shown). Similar results have been obtained for the immunopurified fusion protein, except that the retained fraction is only 5–10 % (Fig. 3 A, B). As a control, β galactosidase did not bind to DNA cellulose and was found in flow through and washes fractions only (Fig 3C). The reason why 90–95 % of the fusion protein was observed in the flow through fraction may be that the renaturation was incomplete after the immunopurification, SDS-PAGE and electroelution steps, or that the actual conformation of the fusion protein impair the availability of the motif involved in DNA binding.

Structure of mod protein

Northern analysis of poly (A⁺) RNA isolated from embryos, with the 1.4 kb EcoR I fragment from *mod* cDNA as hybridization probe, showed that the size of *mod* mRNA is about 2.2 kb (Fig. 4), i. e. 2–300 nucleotides more than the largest cDNA obtained through the immunological screen (page 913).

In order to obtain a full-length cDNA clone, we first probed an EMBO 4 genomic library with the insert of phage LA913. The restriction map and Southern blotting experiments indicated that the genomic clone thus isolated (EMBO 4LA1) contained all the *mod* coding sequence and about 9 kb and 0.5 kb of flanking sequences at the 5' and 3' sides respectively (Fig. 5A). Secondly, the lambda gt 11 cDNA library was screened again using two probes that cover the 5' third and 3' two thirds respectively of the LA913 phage insert. Out of

Figure 5: Restriction maps and sequencing strategy

A. Restriction map of the genomic clone EMBO4 LA1 with the following restriction enzymes: Bam HI $\uparrow$, Cla I $\uparrow$, Eco RI $\downarrow$, HindIII $\uparrow$, Kpn I $\downarrow$, Mlu I $\downarrow$, Sac I $\downarrow$, Xba I $\downarrow$. The boxes represent the cDNA sequence projected on the genomic clone. The 5'-3' polarity is indicated by Asterisk points to the translation start codon.

B. Representation of the five independent cDNA clones primarily isolated with the Mab LA9. The direction of the β -galactosidase fusion is shown on the top.

The direction and extent of sequence determination are shown by horizontal arrows. The 1,4 EcoRI insert was subcloned into EcoRI dephosphorylated bluescript and the Xba I-Sac I (site into the LacZ gene) into Klenow filling in Eco RI dephosphorylated bluescript. Oriented deletions were generated with Exo III and Exo VII according to Yanish-Perron (29).

C. Representation of longest cDNA clones and of the local organization of the genomic clone. Horizontal arrows indicated the sequenced part under each clone.

10⁶ plated phages, 60 gave positive signals. The extremities of the longer cDNA insert of phage LA935 was sequenced. We used the genomic fragment BamHI-MluI lying at the 5' end of the coding sequence to select the longest 5' cDNA. LA954 phage was isolated and the 5' part of the insert was sequenced.

The restriction map and the comparison of cDNA (LA913 and the 5' part of LA954) and genomic (not shown) sequences revealed the existence of four introns in position -333, -255, +244 and +1282.

The nucleotide sequence and the deduced amino-acid sequence are shown in Fig 6. The protein coding sequence begins at the first ATG of the 1632 long ORF, ORF determined by sequencing the Lac Z-cDNA junction. This is supported by the following two observations. Firstly, this ATG is the first start codon preceded in frame by several stop codons in the cDNA sequence (at position -200, -280 ..., Fig. 5). Secondly, the sequence surrounding this ATG codon, CAAAATGG, is in agreement with the consensus, (C/A) AA (A/C)ATGG, of the 'scanning model' for the translation initiation of *Drosophila* messengers (34). At the 3' end, there is a 200 bp-long untranslated sequence before the poly (A⁺) tail. Two AATAAA poly (A) addition signals are present near the 3' end of the untranslated region. The distal signal lays at an appropriate distance (16 bp) upstream of the terminal poly (A) stretch which is found in the LA935 cDNA.

The amino acid sequence derived from the cDNA sequence revealed some interesting features of the protein. The protein is highly hydrophilic with 35% of charged residues (18% acidic, Asp + Glu ; 17% basic, Arg + His + Lys), and 31% of polar residues (including 14% of residues with free hydroxyl groups, Ser + Thr). On the whole, the protein is acidic with a calculated pHi of 4.99. Several regions in the sequence can however be defined on the basis of the local concentration of the charged amino acids (Fig. 7). More than half of the acidic residues are concentrated between position 60 and 134, constituting a highly acidic negatively charged domain that contains 45 Asp and Glu residues. Besides this anionic region two cationic domains are found that contain 35% and 36% of basic residues at the amino and carboxyl termini of the protein. *Mod* protein in *Drosophila* embryo is phosphorylated (Garzino et al unpublished). It is thus noteworthy to mention the presence of a typical cAMP dependent phosphorylation site (35), -RRVS-, in position 327 in the amino acid sequence. Phosphorylation however may occur at other putative sites, particularly at the level of serine or threonine residues lying in an acidic environment. Five potential glycosylation sites exist at positions 213, 268, 426, 441 and 473, but we have no evidence that this type of posttranslational maturation does occur.

Computer searches (Gene Pro, NBRF data banks) with the amino acid sequence of *mod* protein failed to reveal any significant homology with previously sequenced genes or proteins. The acidic domain (amino acid 60 to 134) which contains 60% acidic amino acids (Asp and Glu), presented of course some similarities with the acidic domains of cytoplasmic proteins (myosin, tropomyosin), membrane proteins (sodium channel) or nuclear proteins (myc oncogene, rat glucocorticoid receptor, engrailed and serendipity *Drosophila* proteins). The NH₂ basic domain which contains 23% Lys residues presented some similarities with lysine rich proteins as histone H1. In all cases, we observed between 20 and 30% of similar amino acid on short parts of the sequences.

Mod gene exhibits both maternal and zygotic expression.

In situ hybridization of biotinylated EMBO 4 LA1 phage DNA to salivary gland polytene chromosomes showed that the *mod* gene is located in 100F, the telomeric region of the

[illegible]

right arm of the third chromosome (Fig. 8). No mutation has been described which corresponds to this position. Genomic Southern blot analysis (not shown), performed under high stringency conditions, are also in favor of a unique location of the gene.

In order to analyze the transcription pattern of *mod* gene during development, we have probed dot-blots of staged RNA with the 1.4 kb EcoRI fragment (Fig. 9). The mRNA is abundant in ovaries of adult females, in preblastoderm and early embryos, and detected but at a much lower level during the other developmental stages (late embryos, larvae, pupae and virgin females). Note, however, that the messenger is more abundant in pupae and virgin females than in larvae. The *mod* protein was not detected in preblastoderm embryos, either on Western blots of nuclear proteins or in immunofluorescence assays on embryos sections, and appeared at the cellular blastoderm stage in all the nuclei of the cellular epithelium (21). From these results we deduce that the messengers produced in ovaries by the transcription of *mod* gene are stored in oocytes and preblastoderm embryos and not translated before the blastoderm stage. A posttranscriptional regulation of *mod* gene expression occurs therefore, at least for the maternally inherited mRNA.

DISCUSSION

We have isolated, characterized and sequenced cDNA and genomic clones encoding *mod* antigen, a DNA-binding protein which is differentially expressed during *Drosophila* embryogenesis.

Three independent lines of evidence demonstrate that these clones encode the *mod* protein. First, five independent recombinant phages were screened with Mab LA9 ; inserts of these phages share the same 1.4 kb EcoRI fragment. Second, the apparent size on SDS-PAGE of the cloned protein is appropriate for *mod* polypeptide. Effectively, besides the fusion protein, Mab LA9 recognizes on Western blot a polypeptide that co-migrates with *mod* antigen. The most plausible explanation is that a low level of translation was initiated at the initiation ATG codon of the cDNA insert leading to the production of intact *mod* antigen. The fusion material corresponded rather to translational readthrough of fused LacZ-cDNA sequence. Saxena et al (37) have recently described lyzogens for recombinant phages that also produce both the high molecular weight fusion protein and the low molecular weight polypeptide corresponding to the inserted sequence. Third, polyclonal antibodies raised against the cloned fusion protein also recognize *mod* antigen. Immunofluorescence assays on frozen sections of embryos indicate that antiserum depleted from antibodies against β -galactosidase, stained, as Mab LA9 does, the nuclei of the gut and associated structures. On Western blots of nuclear proteins the depleted antiserum recognizes an unique polypeptide with the same apparent molecular weight as *mod* antigen. In addition the bacterial fusion protein retains the ability noticed for the *Drosophila* protein to bind DNA (Fig. 3). In light of these results we conclude that we have identified the DNA sequence producing the *mod* antigen.

Figure 6: cDNA and predicted protein coding sequence for *Drosophila mod*. The sequence of 2,276 bases of *mod* cDNA was determined by the dideoxy DNA method according to the strategy outlined in Fig. 4. The primary sequence is numbered in the 5' to 3' direction, beginning at the first nucleotide of the ATG initiation codon. The predicted amino acid sequence is shown above the corresponding nucleotide sequence.

Boxed nucleotide sequences indicate, the stop codons in frame with the translation start codon, the surrounding sequence of the start codon and the poly A additive signal. Underlined amino acid sequence indicates a potential phosphorylation site (the serine is boxed). Vertical arrows indicate the position of the introns, determined by restriction mapping and by comparing the sequence data obtained from genomic and cDNA clones (Fig. 4).

Figure 7: Local base concentration of charged residues in the amino acid sequence of *Drosophila mod* protein. Plot of Asp + Glu proportion, computed from position 1 to 544 using an interval of 7 amino acids (upper panel). Percentage of charged residues and the net charge in the two basic and in the acidic regions (lower panel).

The compilation of the nucleotide sequences obtained from three cDNA and from the genomic clone (see Fig. 5) allow us to draw the following conclusions. 1) There is a good size correlation between the mRNA seen in Northern blot (2.2 kb) and the cDNA of Fig. 6 (2261 nt, excluding the poly A tail). 2) This cDNA contains the entire protein coding sequence. The translation starts with the first ATG codon of the 1632 long ORF. This codon is indeed preceded in frame by several stop codons ; moreover, its surrounding sequence is reminiscent of the consensus given by Cavener (34) for translation start. 3) Five exons and four introns have been detected. The 5' untranslated region, that contains two short introns and two exons, appears to be rather long and complex for a *Drosophila* gene. Further experimentation is required for a better understanding of the gene structure, and of its expression.

The ORF encodes a polypeptide of 544 amino acids (60.5 kD). This polypeptide migrates anomalously in SDS-PAGE, with an apparent molecular weight of 78 kD. The discrepancy between apparent and true molecular weights is probably due to an uneven distribution of

Figure 8: Cytogenetic localization of the *mod* gene. Biotinylated EMBO4 LA1 phage DNA was hybridized to squashes of salivary glands chromosomes of third instar larvae and detected with streptavidin-phosphatase. A single hybridization signal was observed at position 100F (arrow) on the right arm of the third chromosome.

charges throughout the protein. The predicted *mod* protein shows an uncommon distribution of charged residues ; four sections can be defined in the amino acid sequence. The NH₂ terminal region (33 residues) is clearly basic with a high proportion of lysine (net charge + 8). A second region, from residue 60 to residue 134, is strongly acidic (net charge -44). The proportion of acidic amino acids is quite unusual consisting of stretches of glutamic and aspartic acids, but neither GAG (poly Glu) repeats as found in HMG proteins (38) nor CAG repeat (39) as found in several *Drosophila* genes encoding nuclear proteins (40). The third region extending from residue 139 to residue 490 is rich in both basic and acidic

Figure 9: Developmental Dot-blot analysis of *mod* transcript. 10 μ g of total RNA from different developmental stages was denatured with formaldehyde and loaded on nitrocellulose filters according to the conditions described by Ait-Ahmed *et al* (1987) (36). Ov, hand dissected ovaries; pbl, preblastoderm embryos (0–1 h 30); bl, blastoderm embryos (2h–3h); g, gastrula embryos (4 h–6 h), l. e., late embryos (12 h–18 h); L, larvae (third instar); P, pupae; ♀, virgin females. Probes were, the 1.4 kb EcoRI fragment purified from LA913 cDNA (A) and an actin genomic fragment as a control (B).

amino acids with a positive net charge of 8. The fourth region is the carboxyl terminus (50 amino acids) of the protein. It is basic, rich in Arg and Lys residues, with a positive net charge of + 17.

Searches for sequence similarities did not reveal anything that would suggest a role for the various domains of the polypeptide. However one can speculate about the possible involvement of charged regions in the interactions that the protein establishes within chromatin. It is tempting to assume that the function of the anionic region is to stimulate the formation of protein-protein complexes and that of carboxyl or amino terminal basic domain to be involved in DNA binding. In order to detect structural motives supporting such interactions we analyzed local secondary structures predicted from amino acid sequence. No DNA-binding typical structure like the homeodomain (41) or zinc finger (42) or the amphipathic helix-loop-amphipathic helix motif (7) were found in the basic domains. Many other proteins known to bind DNA lack such structural motives (43–45). However, short alpha helices were encountered, around position 22 (13 residues), 37 (11 residues), 491–492 (12 residues) and 525–526 (8 residues). These could be candidates for promoting DNA binding. There is no structural paradigm for the acidic domains of nuclear proteins. Activating (acidic) region of transcription factors are rather poorly structured (3, 8) ; the presence of a short amphipathic alpha helix appears however to be an essential element for the activation to proceed (46). The acidic domain of *mod* polypeptide shows a net propensity to adopt an alpha helix conformation with the possibility of an amphipathic helix from residue 85 to 91.

Owing to its modular structure consisting in separate charged domains and to its capacity to bind DNA and protein(s), *mod* protein appears to be a new member of a class of nuclear proteins that are presently receiving much attention with regards to their role in eukaryotic gene regulation, either as transcriptional activators (3, 4) or as structural factors (47). We do not know which cellular processes require the function of the *mod* gene. However, the differential expression of the protein during development, when taken together with the above data obtained from sequence analysis, may suggest a role in the regulation of gene expression during *Drosophila* development.

ACKNOWLEDGMENTS

We thank Ounissa Ait-Ahmed for providing us with dot-blot of staged RNA, Steve Kerridge for determining the cytogenetic localization, Bernard Bellon and Bernard Jacq for their advice in the computer analysis of sequence data, and Steve Kerridge, Raymond Miassod and Roland Rosset for critical reading of the manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Guarente, L. (1988). *Cell* **52**:303–305.
2. Maniatis, T., Goodbourn, S., and Fischer, J.A. (1987). *Science* **236**:1237–1245.
3. Ptashne, M. (1988). *Nature* **335**:683–689.
4. Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. (1988). *Nature* **336**:563–564.
5. Gehring, W.J. and Hiromi, Y. (1986). *A. Rev. Genet.* **20**:147–176.
6. Miller, J., McLachlan, A.D. and Klug, A. (1985). *EMBO J.* **4**:1609–1614.
7. Murre, C., McCaw, P.S., and Baltimore, D. (1989). *Cell* **56**:777–783.
8. Sigler, P. B. (1988). *Nature* **333**:210–212.
9. Lewis, E.B. (1982). In M. Burger (ed.) *Embryonic development : genes and cells*. Alan R. Liss, Inc., New York, pp. 268–288.
10. Scott, M.P. and Carroll, S.B. (1987). *Cell*. **51**:689–698.

11. Levine, L., and Hoey, T. (1988). *Cell* **55**:537–540.
12. Rosenberg, U.B., Schröder, C., Preiss, A., Kienlin, A., Coté, S., Riede, I. and Jäckle, H. (1986). *Nature* **319**:336–339.
13. Vincent, A., Colot, H.V. and Rosbach, H. (1985) *J. Mol. Biol.* **186**: 146–166.
14. Mansukani, A., Crickmore, A., Sherwood, P.W. and Galberg, M.L. (1988). *Mol. Cell. Biol.* **8**:615–623.
15. Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S. and Corces, V.G. (1988). *Genes Dev.* **2**:1205–1215.
16. Poole, S.J., Kauvar, L.H., Drees, B. and Kornberg, T. (1985). *Cell* **40**:37–40.
17. Gehring, W.J. (1987). *Science* **236**:1245–1252.
18. Vincent, A., Kejzlarova-Lepesant, J., Segolat, L., Yanicostas, C. and Lepesant, J.A. (1988). *Mol. Cell. Biol.* **8**:4459–4468.
19. Fasano, L. and Kerridge, K. (1988). *Development* **104**:245–253.
20. O’Kane, K., and Gehring, J.W. (1987). *Proc. Natl. Acad. Sci. USA* **84**:9123–9127.
21. Garzino, V., Moretti, C., and Pradel, J. (1987). *Biol. Cell.* **61**:5–14.
22. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
23. Nolan, J.M., Lee, M.P., Wyckoff, E. and Hsieh, T.S. (1986). *Proc. Natl. Acad. Sci. USA* **83**:3664–3668.
24. Huynh, T.V., Young, R.Y., and Davis, R.W. (1985). In Glover, D.M. Glover (ed), *DNA cloning. A practical Approach*. IRL Press, Oxford, Vol. I, pp. 49–78.
25. Ramsay, G., Stanton, L., Schwab, M. and Bishop, J.M. (1987). *Molec. Cell. Biol.*, **6**:4450–4457.
26. Fyrberg, E.A., Kindle, K.L., and Davidson, N. (1980). *Cell* **19**:365–378.
27. Feinberg, A.P. and Vogelstein, B. (1984) *Anal. Biochem.* **137**:266–267.
28. Engels, W.R., Preston, C.R., Thompson, P. and Eggleston, W.B. (1986). *Focus* **8**:6–8.
29. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). *Gene* **33**:103–119.
30. Vieira, J., and Messing, J. (1987). *Methods in Enzymology*, **153**:3–11.
31. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
32. Sanger, F., and Coulson, A.R. 1978. *FEBS Lett.* **87**:107–110.
33. Bellon, B. (1988). *Nucleic Acids Res.* **16**:1837–1846.
34. Cavener, D.R. (1987). *Nucl. Acids Res.* **15**:1353–1361.
35. Edelman, A.M., Blumenthal, D.K. and Krebs, E.G. (1987). *Ann. Rev. Biochem.* **56**:567–613.
36. Ait-Ahmed, O., Thomas-Cavallin, M., and Rosset, R. (1987). *Dev. Biol.* **122**:153–162.
37. Saxena, A., Padmanabha, R., and Glover, C.V.C. (1987). *Mol. Cell. Biol.* **7**:3409–3417.
38. Goodwin, G.H., and Johns, E.W. (1983). In L.S. Hnilica (ed.) *Chromosomal non histone proteins*, Vol. 3. CRC Press, Inc., pp. 70–87.
39. Schneuwly, S., Kuroiwa, A., Baumgartner, P. and Gehring, W.J. (1986). *EMBO J.* **5**:733–739.
40. Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. and Perrin-Schmitt, F. (1988). *EMBO J.* **7**:2175–2188.
41. Harding, K., Wederen, C., Mc Ginnis, W., and Levine, M. (1985). *Science* **119**:1239–1242.
42. Klug, A., and Rhodes, D. (1987). *T.I.B.S.* **12**:464–469.
43. Pirrota, V., Manet, E., Hardon, E., Bickel, S.E. and Bensen, M. (1987). *EMBO J.* **3**:779–791.
44. Schleif, R. (1988). *Science* **241**:1182–1187.
45. Weigel, D., Jürgens, G., Küttner, F., Seiffert, E. and Jäckle, H. (1989). *Cell* **57**:645–658.
46. Giniger, E., and Ptashne, M. (1987). *Nature* **330**:670–671.
47. Earnshaw, W.C. (1987). *J. Cell. Biol.* **105**:1479–1482.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.