The *period* Clock Locus of *D. melanogaster* Codes for a Proteoglycan

Pranhitha Reddy,† Agnes C. Jacquier,* Nadja Abovich,* Gabriele Petersen,* and Michael Rosbash†
†Department of Biology
Brandeis University
Waltham, Massachusetts 02254
†Department of Biochemistry
Brandeis University
Waltham, Massachusetts 02254

Summary

The period (*per*) gene of *D. melanogaster* is involved in the generation of biological rhythms. The most striking feature of the predicted coding sequence, corresponding to the key 4.5 kb transcript from this locus, is an extensive run of alternating Gly-Thr residues. This is homologous to a series of Gly-Ser repeats in a chondroitin sulfate proteoglycan. To determine whether the *per* transcript codes for a proteoglycan, a region of its coding sequence was expressed (in bacteria) as part of a fusion protein, which was used to immunize rabbits. When the resultant immune sera were used to probe fly protein preparations, they detected an antigen that is present in wild-type flies and absent in a *per* mutant. Biochemical characterization of this antigen indicated that it is indeed a proteoglycan.

Introduction

The best characterized clock gene in *D. melanogaster* is defined by several allelic mutations at the X chromosomal *period* (*per*) locus. Mutations at this locus disrupt, or modulate, biological rhythms operating at widely different time constants (Konopka and Benzer, 1971; Kyriacou and Hall, 1980, 1986). Thiose genetic and behavioral data suggest that the *per* gene product(s) plays a central role in the generation of the fly’s rhythms.

At the molecular level, the mechanism and control of biological clocks are poorly understood. There is no knowledge about the nature and function of the proteins or other molecules that constitute such oscillators. Information on the biochemical nature and structure of the *per* protein should lead to considerable insights into the function of the *per* product in the production or regulation of biological rhythms.

A region of the fly’s genomic DNA has been identified to contain the information required for *per* function (Reddy et al., 1984; Zehring et al., 1984; Hamblen et al., 1986; Bargiello and Young, 1984). In particular, P element-mediated germ-line transformation, using DNA from the *per* locus to restore rhythmicity of mutant flies, has defined the region of DNA required for the biological activity of the *per* gene (Figure 1). In parallel, this region of the genome has been assayed for complementary transcripts. Apparently, a single 4.5 kb RNA is transcribed from this biologically active region. This transcript is polyadenylated (Reddy et al., 1984; Bargiello and Young, 1984) and is enriched in the head of the animal (A. A. James et al., submitted), where the circadian clock may be housed (Konopka et al., 1983). These and other lines of evidence (Figure 1) suggest that this transcript codes for the *per* protein product.

In this study, we present the predicted amino acid sequence of a portion of this protein product as deduced from nucleotide sequencing data. Using recombinant DNA techniques, a small subregion, corresponding to 154 amino acids of the protein, was cloned and expressed in bacteria as a fusion protein. An antibody reagent, developed using this fusion protein, was characterized and used to study a protein product of the 4.5 kb transcript.

Results

Transcript Mapping and Protein Sequence of the 4.5 kb RNA

Our current view of the *per* gene is shown in Figure 1. A single 4.5 kb transcript is detected from this region, although molecular heterogeneity, generated perhaps by differential splicing, has not been excluded. This 4.5 kb RNA is transcribed in embryos, where it is localized to the embryonic nervous system, and in adults where it is enriched in the head (James et al., submitted). We and others have defined, by P element-mediated transformation, the DNA region necessary for biological activity, that is, for the restoration of biological rhythms in *per* mutant hosts. Based on results from two laboratories (Zehring et al., 1984; Hamblen et al., 1986; Bargiello et al., 1984), the minimal region necessary for biological activity consists of approximately 5 kb (open bar in Figure 1); this coincides with most or all of the coding DNA of the 4.5 kb transcript, as defined by cDNA cloning, ribonuclease mapping, and DNA sequencing (Y. Citri, A. C. Jacquier, and H. V. Colot, unpublished data).

The complete sequence and organization of the *per* region will be presented elsewhere. In this communication, we present the DNA sequence of a region of approximately 1.6 kb that contains a single open reading frame. The introns flanking this exon have been defined by ribonuclease mapping and cDNA clone analysis. The single long open reading frame contains, for the most part, codon usage features characteristic of *Drosophila* coding regions (Pustell and Kafatos, 1986). No statistically significant homologies were found from a search of the Protein Sequence Database of the Protein Identification Resource, release 7.0 (November, 1985).

The most striking features of this sequence are highlighted in Figure 2. First, the sequence contains a remarkable run of alternating Gly-Thr residues. This feature is similar to a previously described run of Gly-Ser residues, in which many of the Ser residues constitute the glycosidic attachment sites of the glycosaminoglycan (GAG) side chains to the core protein of this chondroitin sulfate...
Cloning and Expression of an Open Reading Frame of the 4.5 kb Transcript

The 1.5 kb BamHI–HindIII fragment was isolated and sonicated. DNA fragments ranging in size from 350–600 bp were isolated and cloned into the Smal cloning site of the pMR100 expression vector (Gray et al., 1982). Positive clones were identified by their lac' phenotype. Of several such clones identified, two produced high levels of fusion protein. The inserts from these clones were subcloned into pEMBL vectors (Dente et al., 1983) and sequenced to determine the reading frame used in the fusion protein (Sanger et al., 1977). The orientation of the inserts in the pMR100 expression vector was determined by restriction enzyme digestion analysis. One clone, pORFId2, was shown to contain 464 bp, which corresponds to 154 amino acids of the per protein sequence in-frame with β-galactosidase. The amino acid sequence of ORFId2 is indicated in Figure 2 and lies downstream of the Gly-Thr repeat.

The DNA sequences that code for the Gly-Thr repeat and the adjacent downstream ORFId2 have been used as probes in Southern and Northern blots (Figure 3). In a genomic Southern blot analysis, the ORFId2 probe recognizes a single band, as predicted from the genomic restriction map (data not shown). The Gly-Thr probe recognizes two prominent HindIII bands at low stringency, one of which nearly disappears when the blot is washed at normal stringency. The band that remains is of the predicted size from the per restriction map and is not detectable in DNA isolated from flies deleted of the per region (Figure 3B). Similar results were obtained with EcoRI and BamHI restriction digests (data not shown). By Northern blot analysis with adult RNA, the two probes recognize only the 4.5 kb transcript (Figure 3A). From these data, we conclude that the two probes are quite specific for the per region; even the repeated DNA sequence, which codes for the Gly-Thr repeats, has homology to few other regions of the Drosophila genome.

Generation and Characterization of Specific Rabbit Antisera

The pORFId2 fusion protein was isolated from bacterial cell lysates as described (Gray et al., 1986). Briefly, the fusion protein was enriched by using ammonium sulfate fractionation and a β-galactosidase substrate affinity column. In the final purification step, the proteins were fractionated by SDS-PAGE, and the fusion protein was used to immunize rabbits (Gray et al., 1986).

Affinity-purified antisera were assayed for immunoreactivity against the fusion protein using an ELISA assay (Voller et al., 1979) (data not shown) and on Western blots (Towbin et al., 1981; Stahl et al., 1984; and Figure 4A). The Western blots demonstrate that the strain carrying pORFId2 contains a large fusion protein. The differences in size between pORFId2 and pMR200 (λCl-β-galactosidase) fusion proteins is consistent with the predicted insert of 154 amino acids. When probed with affinity-purified antibodies, the lane containing protein from pORFId2 shows a reaction at the expected position, while the lane containing pMR200 protein is blank. From these results we conclude that the affinity-purified antisera con-
Exonic sequences are shown in upper case letters; intervening sequences are shown in lower case. The implied translation of the per transcript, into a sequence of amino acids, is given under the (exonic) DNA sequences. The Gly-Thr repeat and the ORF<sub>2</sub> sequences, defined by sequencing the DNA sequences, the Gly-Thr repeat and the ORF<sub>2</sub> sequences, defined by sequencing the DNA sequences, are indicated between brackets. The Gly-Thr, Thr-Gly, Ser-Gly, and Gly-Ser amino acid pairs are underlined, and N-linked glycosylation sequences are indicated by an asterisk (*). The single-stranded fragment used as a probe in Figure 3 is overlined.
tain antibodies specific for the 154 amino acid insert of pORFId2. Since the antigen used for immunizations was denatured with SDS, we compared the immunoreactivity of the antibody with native protein and with SDS-treated protein. The results shown in Figure 4B indicate that the SDS-treated protein is a significantly better antigen; subsequently, all protein samples were treated with 0.1% SDS prior to reaction with antibody.

Identification of the per Products in Fly Head Protein Extracts

Preliminary experiments, using Western blot techniques (Towbin et al., 1981) on fly protein extracts fractionated on SDS-PAGE, failed to show any strong and consistent reactivity of the antibodies with fly proteins. This was interpreted to indicate that the nature and/or abundance of the per protein(s) preclude its detection with this technique. In fact, proteoglycans are notoriously difficult to assay by protein blotting (Hascall and Kimura, 1982). Thus, further experiments, designed to determine whether this per-encoded antigen is a proteoglycan, were performed. Such proteins are identified and isolated by their very unusual biochemical characteristics, which have been reviewed by Hascall and Kimura (1982). The sulfate groups on their GAG side chains give these molecules a high net negative charge, responsible for their characteristic behavior on anionic column chromatography. The GAG side chains also confer upon these molecules a very large hydrodynamic volume, which can be used to identify and isolate these molecules by molecular sieve chromatography.

Following established procedures for proteoglycan extraction (Hascall and Kimura, 1982; Yanagishita et al., 1988), heads from wild-type and per- (Df(1)TEM202/Df(1)64j4) flies were assayed. Df(1)TEM202/Df(1)64j4 females are deleted of at least 10 kb of per-region DNA (Figure 1), are behaviorally arrhythmic (for example see Hamblen et al., 1988), and are missing the 4.5 kb transcript (Reddy et al., 1984); they provide us with an important negative control. Heads were chosen for this initial investigation, since the 4.5 kb RNA is enriched in this region of the animal (James et al., submitted). The protein extracts were desalted and then rechromatographed on a DEAE-Sephacel column, using step-wise elution. Aliquots of the eluted material were analyzed on DEAE membranes using immunoblot procedures (Towbin et al., 1981; Stahl et al., 1984). Results of a representative experiment are shown in Figure 5A. There is a striking difference between the elution profile of antigen from wild-type (CS) versus per- flies (Df/Df) at 1.0 M NaCl; no antigenic material elutes at 1.0 M NaCl in the protein from per- flies. The nature of the cross-reacting antigen present in the 150 mM eluate of both deletion and wild-type fly protein was not studied further and is discussed below. To characterize

Figure 3. Northern and Southern Blot Analysis from the GT Repeat and ORFId2

(A) Approximately 2–4 µg of poly(A)+ RNA from adult flies was electrophoresed on a formaldehyde agarose gel and hybridized to the following probes using Northern blot technique (Reddy et al., 1984); lane 1 was hybridized to a single-strand DNA probe synthesized from ORFId2 cloned in pEMBL 18+; lane 2 was hybridized with a single-strand DNA probe from the Gly/Thr repeat region indicated in Figure 2. The position of the 4.5 kb RNA is indicated by the arrows. We previously identified an approximately 1.0 kb RNA from this region of the per locus (Reddy et al., 1984). The use of very pure single-stranded probes (as detailed in Experimental Procedures) has subsequently revealed only the 4.5 kb RNA, so we assume the smaller RNA to have been an artifact. (B) Approximately 10 µg of genomic DNA from adult flies was digested with the restriction enzyme HindIII (20 U, Boehringer Mannheim), electrophoresed on agarose gels, and hybridized to the single-strand DNA probe from the GT region (Figure 2) using Southern blot techniques (Reddy et al., 1984). Lanes 1 and 3 contain DNA from wild-type (Canton-S) flies, and lanes 2 and 4 contain DNA from per-
Further the nature of the per-specific antigens, the 1.0 M NaCl fraction was diluted and rechromatographed on a DEAE-Sephacel column using a gradient elution. The results of this experiment are shown in Figure 5B and demonstrate that the antigen elutes as a single peak at a salt concentration of approximately 0.75-0.77 M NaCl. Similar elution profiles (ranging from 0.75-0.80 M NaCl) were observed in two other experiments. This elution profile is very characteristic of proteoglycans, which elute from DEAE columns at salt concentrations at or above 0.5 M NaCl (Yanagishita et al., 1986). Glycoproteins without sulfate residues elute at much lower salt concentrations (Yanagishita et al., 1986).

Molecular Sieve Chromatography
To investigate further the nature of the wild-type-specific antigen that elutes from DEAE-Sephacel at 1 M NaCl, this antigen was rechromatographed on Sepharose-CL-6B. The results (Figure 6) show that most of the antigenic material is present in the excluded volume, indicating an apparent molecular weight of one million daltons or greater. The antigen retains this elution profile when a stronger denaturant, 4 M guanidine hydrochloride, is substituted for urea in the buffer used for the CL-6B-Sepharose chromatography (data not shown).

Enzymatic Treatment of the Fly Antigen
Chemical and enzymatic treatments that specifically remove the GAG side chains from proteoglycans are widely used to investigate the structure of these molecules (Selvin et al., 1985). As predicted, these treatments radically change all the properties of the proteoglycan that were initially used to characterize it. To verify further the proteoglycan nature of the per product, we subjected the per-specific antigenic material (the 1.0 M NaCl eluate from the DEAE column) to enzymatic treatments. We used two enzymes, heparitin sulfate lyase (heparitinase) and chondroitinase ABC; heparitinase specifically cleaves heparan sulfate linkages and chondroitinase ABC is specific for chondroitin/dermatan sulfate linkages. Figure 7 shows the effect of heparitinase treatment on the elution profile of the per antigen on DEAE and CL-6B columns. Incubation of the antigen under digestion conditions, in the absence of enzyme, does not detectably change its elution profile on either DEAE-Sephacel or Sepharose-CL-6B columns. After heparitinase treatment, the antigen elutes with a heterogeneous profile well below 0.75 M. This material behaves similarly on the Sepharose-CL-6B column in that it chromatographs as a mixture of proteins much smaller than the untreated material. Similar experiments using chon-
droitinase ABC show that this enzyme has no effect on the properties of the antigen (data not shown). These effects in the elution profile and the heterogeneity of the elution profile subsequent to heparitinase treatment could be due to one or a combination of the following reasons: incomplete digestion of the heparan sulfate side chains; the presence of other GAG side chains (which are almost certainly present, since high salt concentrations, i.e., 0.58 M-0.08 M NaCl, are still required to elute the heparitinase-treated antigen); other heterogeneous types of protein modification and glycosylation; and proteolytic cleavage of the antigen by contaminating proteases in the heparitin sulfate lyase preparation. Nevertheless, the results of this experiment are consistent with the results shown above and suggest that the per locus codes for a proteoglycan of the heparan/heparin sulfate variety. This assignment is also consistent with the elution position from DEAE-Sephacel, which indicates oversulfation, as is characteristic of heparan/heparin sulfate proteoglycans (V. C. Hascall, personal communication).

Discussion

The repeating DNA sequence, which has been assigned a coding function for a Gly-Thr repeat, lies in the middle of a long open reading frame. All of its properties, including its presence in the middle of the coding region of a number of cDNA clones, indicate that its function is to code for the indicated amino acid sequence. Moreover, inspection of third base changes in this sequence indicates that the DNA repeat is inexact but maintains an exact protein repeat (for 20 pairs).

We undertook the immunological side of this study to investigate whether the Gly-Thr repeat may be an attachment site for GAG side chains as suggested by the homology to the Gly-Ser repeats of a chondroitin sulfate core protein (Bourdon et al., 1985). While this homology is impressive and statistically significant (Lipman and Pearson, 1985) and while several cellular (heparin/heparan sulfate and chondroitin/dermatan sulfate) proteoglycans are distinguished by protein cores that contain primarily a copolymer of Ser and Gly (Metcalfe et al., 1980; Robinson et al., 1978), not a single case has been reported of such a proteoglycan with a Gly-Thr repeat. Only the keratan sulfate side chains of cartilage proteoglycans may be attached to threonine as well as serine. However, these proteins do not seem to have the amino acid repeat observed in the other cases cited above (Seno et al., 1985; Bray et al., 1967). Since the threonines in the per protein might also serve as attachment sites for O-linked oligosaccharides rather than GAG side chains, we set out to characterize the nature of the macromolecule coded for by this region of the per DNA. The data presented make a strong case that the per DNA codes, at least in part, for...
per Gene in Drosophila Codes for a Proteoglycan

Figure 7. Heparitinase Treatment of Fly Antigen

The high salt DEAE fraction, prepared as described in Experimental Procedures, was incubated with or without heparitin sulfate lyase as recommended by the manufacturers. Each of these samples was split in half. One half was adjusted to 0.2 M NaCl in urea buffer and chromatographed on a 0.1 ml DEAE-Sephacel column, using a linear gradient of 0.2-1.0 M NaCl in 5 ml of urea buffer. The other half was adjusted to 4 M guanidine hydrochloride, 0.5% Triton X-100, 50 mM sodium acetate (pH 6.0) and was chromatographed on a Sepharose-CL66 column (14 x 0.5 cm) in the same buffer. Fractions from the DEAE-Sephacel columns were diluted 5-fold in urea buffer, and fractions from Sepharose-CL66 columns were diluted 20-fold into 50 mM sodium acetate (pH 6.0), 0.2% Triton X-100 buffer and dotted onto DEAE membranes. The dot blots were hybridized to affinity-purified antibody using Western blot methods as in Figure 5. The autoradiograms were scanned on a densitometer, and the relative optical densities of the dots (in the linear range) for each fraction are shown. (A) The DEAE-Sephacel chromatography; (B) the Sepharose-CL66 chromatography. (x——x) heparitin sulfate lyase treatments; (——) mock inoculation controls; (——-) NaCl gradient.

a proteoglycan. If the threonines present in the Gly-Thr repeat are the attachment sites for GAG side chains, it is somewhat surprising that the small DNA probe containing this region manifests such specificity for the per region. The fact that only one other band is clearly detectable on Southern blots at low stringency suggests that there are relatively few other Gly-Thr repeats of substantial length in the Drosophila genome (Figure 3). Perhaps most other Drosophila proteoglycan genes employ Gly-Ser repeats for GAG chain attachment.

Though we have demonstrated that at least some of the protein (or proteins) with which the antisera react is proteoglycan-like, we have not proven that the Gly-Thr repeat is an attachment site for proteoglycan side chains. As there are ample Gly-Ser pairs sprinkled throughout this (Figure 2) and other regions (data not shown) of the per gene, it is possible that these serines are the only attachment sites for proteoglycan side chains. The importance of the Gly-Thr repeat for the biochemical properties of the proteoglycan molecules reported here, and the biological properties of expression of per, will be tested by recombinant DNA techniques and P element-mediated transformation into the Drosophila germ line.

There is also a considerable amount of antigenic activity in the 0.15 M eluate of the DEAE-Sephacel column when total protein from fly heads is assayed. Most of this antigenic activity is likely to be artifactual, as these fractions react with preimmune sera, and rather similar amounts of antigenic material are found in protein from per− flies (Figure 5A). However, some of this antigenic material may be genuine (i.e., only present in wild-type fly preparations), indicating that some ORF1d2 antigen may be located in nonproteoglycan protein. Antigenic material of this nature could be due to inadvertent (artifactual) proteolysis; an antigen lacking glycosaminoglycan side chains; differential glycosylation; or perhaps a protein that is missing GAG attachment sites. This latter possibility could be the product of a specific proteolytic event separating the antigenic region from the Gly-Thr repeat or an event at the nucleic acid level that generates an mRNA containing the antigen-coding region but missing the Gly-Thr repeat.

In this vein, it is not proven that the 4.5 kb per DNA is a unique entity, nor that it codes for the clock protein. While the coincidence of the biologically active DNA and the coding DNA of the 4.5 kb transcript makes this a likely possibility, definitive proof is lacking. Other rare transcripts from this region and/or differential splicing have not been excluded. Hence, it is possible that the proteoglycan characterized in this report is not the correct or the only per-clock protein encoded within the per locus. Despite these caveats, the 4.5 transcript remains the best candidate for the clock protein mRNA. Since the Gly-Thr repeat is coded for by this transcript, and as the biochemical properties of a protein(s) coded for by this region indicate that it is a proteoglycan, our working hypothesis is that a principal product of the per gene is a proteoglycan and that this material is related to the biochemistry of clock function. Tissue localization studies indicate that the 4.5 kb transcript is predominantly or exclusively present in the nervous system (James et al., submitted). Therefore, we predict that this per product is a nervous system-specific proteoglycan. Studies using crude membrane preparations from fly heads indicate that the antigenic material is associated with these membranes (P.
Reddy, unpublished data). The exact nature of this association will be investigated.

Proteoglycans have been recognized in a variety of nervous system-related studies. They are found in many locations (e.g., intracellular, plasma membrane-associated, extracellular matrix) and have been ascribed a variety of functions including, but not limited to, roles in development, synaptogenesis, maintenance of synapses, secretion, and intercellular communication (Anderson and Fambrough, 1983; Carlson and Kelly, 1983; Brandon et al., 1985; Matthew et al., 1985). Although it is premature to speculate further about the relationship between proteoglycans and clock function, future experiments should substantiate the validity of this connection and perhaps shed some light on the biochemical role that proteoglycans play in the generation or maintenance of biological rhythms.

Experimental Procedures

DNA Sequencing of the per Gene

The 8.0 kb EcoRI fragment containing most of the 4.5 kb RNA was sequenced using the dideoxy method (Sanger et al., 1977). Fragments suitable for sequencing were generated from the 8.0 kb fragment cloned in the pEMBL vectors (Dente et al., 1983) by creating an ordered set of deletions using Exonuclease III by the method described by Henikoff (1984). Both strands of the DNA were sequenced.

Probes for Northern Blot Analysis

Single-stranded probes were used for Northern blot analysis generated by transcribing single-strand DNA from a small subclone containing the indicated insert (GT repeat region or ORFId2 in Figure 3) in pCMV-10+ (Dente et al., 1983). The sequencing primer (NEB) was used to transcribe the insert sequence. After transcription (with radio labeled dATP and dCTP) the insert was excised by restriction enzyme digestion and denaturation, and was purified as a radioactive, single-strand fragment from a sequencing gel.

Preparation and Assay of Antibodies Specific for the Inserted Polypeptide Sequence of pORFId2

The rabbit antisera were assayed for immunoreactivity against the fusion protein using ELISA assays and Western blots. Antibodies directed specifically against the per sequences were purified using affinity chromatography (Gray et al., 1986). The antisera were first adsorbed onto an affinity column, consisting of affigel-10 linked to a protein preparation enriched for pORFId2 fusion protein. This adsorbed antibody fraction was eluted and rebound to a second column consisting of heparin sulfate lyase (Miles Lab) or without heparitin sulfate lyase (Miles Lab) or without heparitin sulfate lyase for 1 hr at 42°C in the buffer recommended by the manufacturers.

Acknowledgments

We thank V. C. Hascall (National Institutes of Health, Bethesda, Maryland) for his invaluable advice concerning the proteoglycan analysis and for his assistance in the preparation of this manuscript, and J. Spiegel, J. Sue, and E. Rosen for reagents plus helpful suggestions on the immunological studies. We are grateful to J. Hall and H. V. Colot for their critical reading of the manuscript. M. R. also thanks W. Pearson for advice on the use of the DFASTP computer program and R. Hynes for pointing out the publication on the f Syrians' repeats of the rat chondroitin sulfate proteoglycan. This work was supported by a National Institutes of Health grant (PGM 33266) to M. R.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 2, 1986; revised April 23, 1986.

References


