

Translation of *Drosophila melanogaster* sequences in *Escherichia coli*

(recombinant DNA/plasmids/protein synthesis/minicells/gel electrophoresis)

ALAIN RAMBACH* AND DAVID S. HOGNESS†

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Contributed by David S. Hogness, August 22, 1977

ABSTRACT Thirty-seven independently cloned segments of *Drosophila melanogaster* DNA (Dm segments) were individually tested for their ability to promote the synthesis of new polypeptides in *Escherichia coli* K-12. The cloning vector was the pSC101 plasmid and the test system consisted of *E. coli* K-12 minicells that contained the hybrid pDm plasmids. Each of four pDm plasmids produced a new polypeptide, and one, pDm107, was selected for detailed mapping of the sequences required for the translation of its 38,000-dalton polypeptide, the Dm107 protein. Mapping was accomplished by constructing (i) deletion derivatives of pDm107 and (ii) new plasmids consisting of fragments of the Dm107 segment inserted into other vectors, and then testing these hybrids for their ability to promote the synthesis of the Dm107 protein, or truncated versions of this protein, in minicells. The 1000 base pairs of sequences that are translated to yield the Dm107 protein were thereby mapped at the center of the 18,000-base pair Dm107 segment, which consists of nonrepetitive sequences located at the base of the right arm of chromosome 2. The four polypeptides produced by the four pDm plasmids require sequences of 4000 base pairs for their translation, and the total amount of DNA in the 37 cloned Dm segments that were tested is approximately 400,000 base pairs. Because no new polypeptides were detected with the remaining 33 pDm plasmids, the fraction of *D. melanogaster* sequences that can be efficiently translated in *E. coli* K-12 is estimated to be 1×10^{-2} .

Hybrid DNA molecules consisting of *Drosophila melanogaster* DNA segments (Dm segments) inserted into bacterial plasmids have been constructed *in vitro* and cloned by propagation in *Escherichia coli* K-12 (1, 2). Primary interest in these cloned segments has centered on the unparalleled advantages they offer for the molecular analysis of gene structure, organization, and expression in *D. melanogaster* chromosomes (1-7). By contrast, only scant attention has been paid to the basic question of whether structural genes carried by the cloned Dm segments can be expressed in the bacterial host. It is this question that we address here.

In the first part of this paper we describe experiments in which 37 independently cloned Dm segments were screened for their ability to promote the synthesis of new polypeptides in *E. coli* K-12. These segments had been inserted into the tetracycline resistance plasmid pSC101, to form hybrids that are called pDm plasmids. The pDm clones were unselected because one of our purposes was to obtain an estimate of the fraction of *D. melanogaster* sequences that can be translated in *E. coli*. Only four new polypeptides were observed as a result of the insertion of a total of approximately 400,000 base pairs (400 kb) of *D. melanogaster* DNA among the 37 pDm plasmids. One of the four plasmids that produced a new polypeptide was ex-

amined in detail to identify the sequences required for its synthesis. The mapping of these sequences within the Dm segment constitutes the second part of this paper.

MATERIALS AND METHODS

E. coli K-12 strains HB101 (8) and the minicell-producing P678-54 (9) have been described. AR1062 is a restrictionless (HsdR⁻) derivative of P678-54 that we made by phage P1 transduction. The plasmids pSC101, pSC105 (10), and pDm101-to-107 (2) have been described. Plasmids pDm800 to pDm831 were formed by the insertion of randomly sheared *D. melanogaster* (Dm) segments into pSC101 at its *EcoRI* site according to the dA-dT connector method (1). Transformations were carried out as described by Glover (11). All plasmids containing Dm segments were propagated under EK1 and P2 containment conditions, as defined by the "National Institutes of Health Recombinant DNA Research Guidelines" (*Fed. Reg.*, July 7, 1976).

Restriction endonucleases *EcoRI* (12), *BamHI* (13), *HindIII* (14), and *Sst I* (S. Goff and A. Rambach, unpublished isolation from *Streptomyces stanfordi*, ATCC 29415) were provided by S. Goff and prepared according to the indicated references. Lengths of DNA restriction fragments were obtained by agarose gel electrophoresis (5). Plasmid DNAs were isolated as described by Wensink *et al.* (1), except that the following simpler procedure was used for the analysis of the pDm107 derivatives (Fig. 2). Bacterial lysates were obtained as before (1), except that 0.05% Triton X-100 was used in place of 0.5% Brij 58. The lysates were heated to 65° for 15 min, the white aggregates formed by the heating were removed by centrifugation (Sorvall SS 34 rotor, 10,000 rpm, 10 min), and 2 ml of the supernatant was made 0.25 M NaCl and 10% in polyethyleneglycol (Carbowax 6000) by addition of 5 M NaCl and 40% polyethyleneglycol. After overnight storage at 4°, centrifugation (Sorvall SS 34 rotor, 5,000 rpm, 10 min) yielded a pellet that was suspended in 1 ml of 0.25 M NaCl/1.0 mM EDTA/10 mM Tris-HCl, pH 8.0, and then mixed with 2 volumes of ethanol. After overnight storage at -20°, the mixture was centrifuged (Sorvall SS 12 rotor, 10,000 rpm, 10 min) and the pellet was suspended in 0.5 ml of 1.0 mM EDTA/10 mM Tris-HCl, pH 8.0 to yield a preparation that is sufficiently enriched in plasmid DNA to allow restriction analysis and transformation. This method has the advantage that it allows one to screen large numbers of clones.

Abbreviations: kb, 1000 bases or base pairs in single- or double-stranded nucleic acids; NaDodSO₄, sodium dodecyl sulfate; Dm, derived from *Drosophila melanogaster*; Tc^R and Tc^S, tetracycline resistant and sensitive, respectively; Km^R, kanamycin resistant.

* Present address: Institut Pasteur, 25 rue du Dr. Roux, 75015 Paris, France.

† To whom reprint requests should be addressed.

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. §1734 solely to indicate this fact.

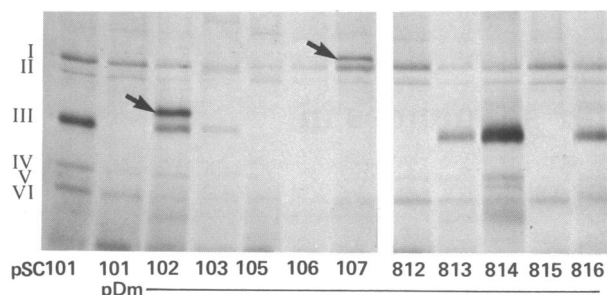


FIG. 1. Electrophoretic patterns of the polypeptides synthesized in minicells containing pSC101 or pDm plasmids. A fresh colony of AR1062 (\pm plasmid) was suspended in 25 ml of M9 medium (16) supplemented to 0.5% glucose, 1% vitamin-free casamino acids (Difco), and, for tetracycline-resistant (Tc^R) bacteria, tetracycline at 15 μ g/ml. After 16 hr of growth with agitation at 37 $^\circ$, bacteria were harvested by centrifugation (Sorvall SS 34 rotor, 10,000 rpm, 10 min) and re-suspended in 0.5 ml of BSG (0.85% NaCl/0.03% KH_2PO_4 /0.06% Na_2HPO_4 /gelatin at 100 μ g/ml). Minicells were then purified at room temperature by the following modification of the technique of Roozen *et al.* (15). Zone sedimentation of the bacterial suspension through 4.0 ml of a 10–30% sucrose gradient in BSG (Sorvall GLC-1 rotor, 3,000 rpm, 25 min) produced a band of minicells that was removed by a pasteur pipette, harvested by centrifugation (Sorvall SS 12 rotor, 10,000 rpm, 10 min), resuspended in 0.5 ml of BSG, and purified through a second sucrose gradient. The purified minicells (≤ 1 bacterium/ 10^7 minicells) were harvested as before, resuspended in 0.75 ml of M9/0.5% glucose plus 0.25 ml of Difco methionine assay medium, and labeled by the addition of 50 μ Ci of [^{35}S]methionine (250 Ci/mmol) and incubation of the cells for 1 hr at 37 $^\circ$. The labeled minicells were harvested at 4 $^\circ$ as before, washed twice in 2 ml of M9, resuspended in 0.2 ml of lysis buffer (2% NaDodSO $_4$ /10 mM dithiothreitol/62.5 mM Tris-HCl, pH 6.8), and boiled for 3 min. After addition of 1 ml of 80% (vol/vol) acetone and incubation at -20 $^\circ$ for 15 min, the precipitated protein was collected by centrifugation, washed two times by suspension of 1 ml of 80% acetone, and dissolved in 50 μ l of lysis buffer containing 20% sucrose and a dash of bromophenol blue by boiling for 3 min. Samples (20 μ l) were layered onto a slab gel (13% acrylamide:*N,N*'-methylenebisacrylamide at 30:0.8/0.28 M Tris-HCl, pH 8.8/0.1% NaDodSO $_4$) and electrophoresed at 12 V/cm for 3–4 hr at room temperature. Autoradiographs of the dried gels were developed after 2–7 days of exposure. Apparent molecular weights for the six bands produced by pSC101 (I, 36,000; II, 34,000; III, 26,000; IV, 22,000; V, 20,500; VI, 20,000) were assigned on the basis of their mobilities relative to those for the light (25,000) and heavy (55,000) chains of hamster gamma globulin. The apparent molecular weights for the new bands produced by certain pDm plasmids (arrows; see *text*) were similarly derived from the values for the pSC101 bands.

RESULTS

The Test System. The 37 pDm plasmids divide into two groups according to the method used to insert the Dm segment at the *EcoRI* site of pSC101. In six (pDm101 to 103 and pDm105 to 107) the connection between the Dm and pSC101 DNAs was effected by ligation of *EcoRI* termini (2); in the remaining 31 (pDm800 to 803 and pDm805 to 831), the connection consisted of dA-dT joints (1).

Detection of the set of polypeptides specified by a given plasmid is difficult because the host polypeptides far outweigh those produced by the plasmid. We therefore turned to the minicell technique for measuring plasmid proteins (15). Minicells are produced by budding during the growth of certain *E. coli* K-12 mutants. They do not contain the bacterial chromosome, but do contain the plasmids carried by the parental cells. The synthesis of host proteins in minicells decreases dramatically due to the decay of mRNA that cannot be regenerated, whereas plasmid protein synthesis is maintained. The 37 pDm plasmids were therefore transferred to a restrictionless, minicell-producing strain, AR1062, by transformation. Mini-

cells containing the desired plasmids were then isolated and labeled with [^{35}S]methionine; their polypeptides were separated by sodium dodecyl sulfate (NaDodSO $_4$) gel electrophoresis, and the labeled polypeptides in the gel were detected by autoradiography (see legend, Fig. 1). The test for a Dm-dependent polypeptide depends on the observation of a new band in the autoradiographic pattern generated by a given pDm that is not present in the pattern generated by pSC101.

The Screen. The pattern generated by pSC101 is characterized by the six bands that are labeled I–VI in Fig. 1. After comparable exposure, the autoradiographs generated by control minicells that lack plasmids are either completely blank or contain faint bands that are also occasionally visible in autoradiographs from plasmid-containing minicells (not shown). None of the six pSC101 bands can be detected in the controls. The pSC101 pattern given in Fig. 1 is like that observed by others (17–19), although different groups report somewhat different molecular weights for the pSC101 polypeptides, as estimated from their mobilities in the NaDodSO $_4$ gels. The apparent molecular weights that we observe (legend, Fig. 1) correspond closely to those given by Meagher *et al.* (19), except that they report a small polypeptide of 14,000 daltons. We do not score autoradiographic responses in the region of the gels corresponding to this and lower molecular weight polypeptides because such responses are variable, diffuse, and of dubious significance, a condition that Meagher *et al.* (19) have also noted.

The bands produced by 33 of the 37 pDm plasmids can all be accounted for by the pSC101 bands or by the faint bands of the controls (e.g., pDm101, 103–106, 812–816; Fig. 1). The four remaining hybrids each exhibit a single, well-defined band that is not produced by pSC101 or the controls. Those obtained with pDm102 and 107 are indicated by the arrows in Fig. 1 and have apparent molecular weights of 29,000 and 38,000, respectively. The new bands formed by pDm820 and 829 (not shown) exhibit apparent molecular weights of 23,000 and 50,000. All but pDm820 were retested with a second preparation of minicells, and each reproduced the original result. In addition, minicells from two new clones of AR1062 [pDm107], obtained from a second purified preparation of pDm107 DNA, were examined and found to give the same autoradiographic pattern as that shown in Fig. 1. While five of the six pSC101 bands can generally be detected in the pDm autoradiographs, band III, which is the most variable in relative intensity, was not detected, or was barely detectable in 15 of the 33 pDm patterns that do not contain an extra band, and in one of the four that do.

Mapping the Sequence Required for the Synthesis of the Dm107 Protein. Among the four hybrids that generate a new band, the sequences in two (pDm102 and 107) had previously been characterized with respect to repetition frequency and location in the genome (2). Of these, pDm107 was chosen for further study because it promotes the synthesis of the larger polypeptide, which we chose to call the Dm107 protein for reasons that will become apparent. The Dm107 segment is 18 kb in length and consists of nonrepetitive sequences that are located at a single site in the right arm of chromosome 2 (region 41D).

To map the sequences required for the synthesis of the Dm107 protein, the cleavage sites for four restriction endonucleases were first located in the circular pDm107 DNA. The resulting map is shown at the top of Fig. 2. Deletion derivatives of pDm107 that lack one or more of the *Bam* I, *Hind*III, or *Sst* I restriction fragments were then constructed and cloned in HB101, as described in the legend to Fig. 2. Diagrams for six of these deletions are given below the pDm107 map. Each of

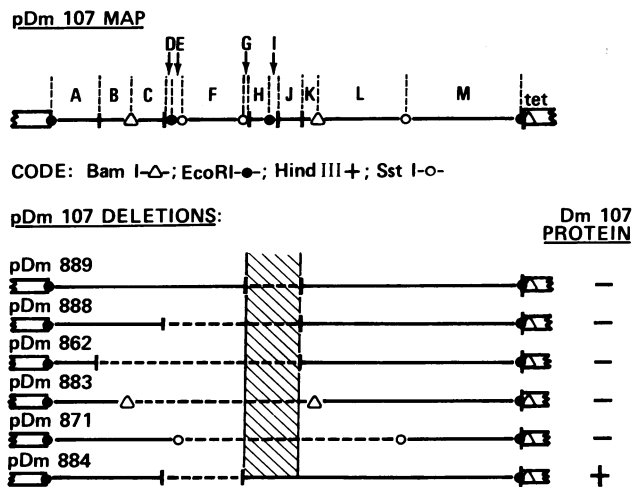


FIG. 2. Map of pDm107 and the capacity of its deletion mutants to promote the synthesis of the Dm107 protein. The lettered segments between cleavage sites in Dm107 have the following approximate lengths, in kb units: A, 1.9; B, 1.2; C, 1.4; D, 0.2; E, 0.4; F, 2.4; G, 0.2; H, 0.8; I, 0.4; J, 0.9; K, 0.6; L, 3.4; M, 4.5. The pSC101 DNA is indicated by the open bars; in addition to the *EcoRI* joints, it contains single *HindIII* and *Bam I* sites that are 0.03 kb and 0.3 kb from the right-hand *EcoRI* joint. The gene(s) responsible for tetracycline resistance (*tet*) are located in this region (see text). The pDm862, 883, and 871 deletions (horizontal broken lines) were constructed by method A: pDm107 DNA was digested to completion with *HindIII*, *Bam I*, and *Sst I*, respectively, diluted to 20 μg of DNA per ml, and treated with *E. coli* DNA ligase at 0.1 $\mu\text{g}/\text{ml}$ for 18 hr at 10° in the standard solvent (5). The supercoiled DNAs larger than pSC101 but smaller than pDm107 were fractionated by electrophoresis in 0.4% agarose gels and used to transform HB101 to Tc^R . Transformants were screened for the desired deletions by partial purification and subsequent restriction fragment analysis of their plasmid DNAs. The pDm884, 888, and 889 deletions were constructed by method B: pDm107 DNA was only partially digested with *HindIII* (as determined by agarose gel electrophoresis of the fragments), and ligation, cloning, and identification of the deletions were then carried out as in method A except that the DNA concentration during ligation was decreased to 1.5 $\mu\text{g}/\text{ml}$, and the products were not fractionated by gel electrophoresis prior to transformation.

these deletion mutants was transferred to AR1062, and autoradiographs of the labeled polypeptides were prepared from the respective minicells in the standard manner. The right-hand column in Fig. 2 indicates the failure (-) or success (+) of each deletion in promoting the synthesis of the Dm107 protein.

All of the deletions that lack the H-J region of the Dm107 DNA are defective for this synthesis; hence, we conclude that part of all of the required sequences are located here. The H-J segment is 2.1 kb long, about twice the length required to code for a polypeptide of 38,000 daltons. pDm884 is not defective for Dm107 protein synthesis, and, because it lacks the adjacent D-G segment, none of the required sequences can be located here.

Additional mapping data were obtained from the synthetic capacities of hybrid plasmids consisting of the C-K *Bam I* fragment of Dm107 inserted into another plasmid vector, pSC105 (10). Fig. 3 shows that pSC105 consists of two *EcoRI* fragments: one is the pSC101 DNA, and the other contains a *kan* gene that confers kanamycin resistance (Km^R) to its hosts. Cells carrying pSC105 are therefore Tc^RKm^R . This is a useful vector because (i) it contains a single *Bam I* site, and (ii) insertion at this locus yields hybrids that produce Tc^SKm^R transformants, evidently by inactivation of a *tet* gene. The desired hybrids were therefore obtained by screening Tc^SKm^R transformants obtained from a ligated mixture of *Bam I* digests

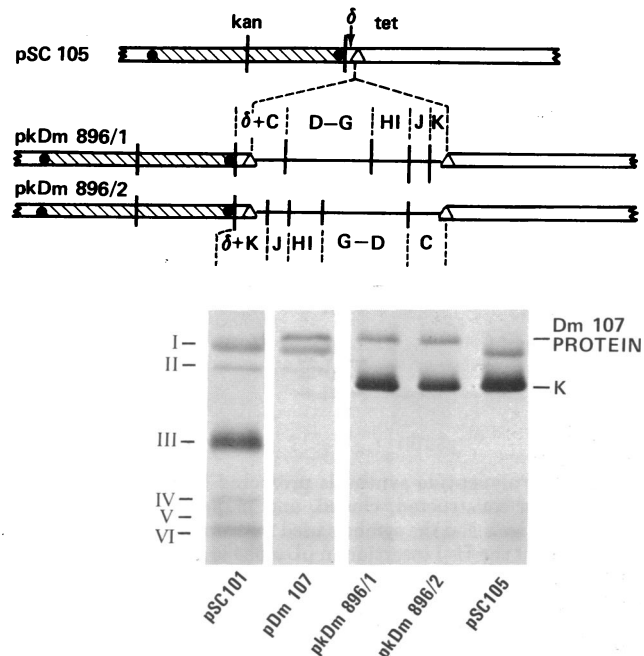


FIG. 3. Orientation and polypeptide synthesis of pSC105 hybrids containing the C-K fragment of Dm107. The symbols used in Fig. 2 apply here. The hatched part of the bar represents the 7.2-kb *kan* fragment (5) that was inserted at the *EcoRI* site of pSC101 to create pSC105 (10). The C-K *Bam I* fragment was inserted into pSC105 by mixing *Bam I* digests of pSC105 (10 $\mu\text{g}/\text{ml}$) and pDm107 (60 $\mu\text{g}/\text{ml}$) DNAs, and ligating this mixture as described in the legend to Fig. 2. Among 108 Km^R clones obtained by transformation of HB101 with the ligated DNA, 19 were Tc^S , and of these, 9 contained plasmids consisting of the C-K *Bam I* fragment inserted into pSC105, as shown by electrophoresis of *Bam I* digests of the 19 plasmid DNAs. The 9 plasmids divide into the two orientation classes represented by pkDm896/1 and pkDm896/2 (see text). The strong K band produced by pSC105 and its derivatives appears to result from the *kan* fragment because we have observed it with pML2, a plasmid consisting of the *kan* fragment inserted into the Cole1 plasmid (20). The insertion of the C-K fragment at the *Bam I* site of pSC105 to produce the two pkDm derivatives can be seen to result in the loss of the pSC101 band I. Because this insertion also results in a Tc^R -to- Tc^S conversion, we suppose that band I derives from a *tet* gene. All of the other six pSC101 bands, except the variable band III, are formed by all three plasmids.

of pDm107 and pSC105 DNAs (Fig. 3 legend). Nine independently cloned hybrids consisting of the C-K *Bam I* fragment of Dm107 inserted into pSC105 were obtained in this manner. They divide into two classes: three have the orientation indicated for pkDm896/1 in Fig. 3, and six have the opposite orientation exhibited by pkDm896/2. The two classes were distinguished by electrophoresis of their *HindIII* fragments in 0.5% agarose gels, where the length of the $\delta+C$ fragment produced by pkDm896/1 is easily distinguished from that of the $\delta+K$ fragment of pkDm896/2 (Fig. 3).

pkDm896/1 and pkDm896/2 were transferred to AR1062, as was pSC105. Autoradiographs of the labeled polypeptides formed in the respective minicells are given in Fig. 3. Both of the pkDm plasmids reproducibly promote the synthesis of the Dm107 protein, indicating that all of the sequences required for this synthesis are contained within the C-K segment of Dm107. Because we have previously shown that sequences within the D-G segment are not required, those that are can be further localized to the C and H-K segments. This result is consistent with the previous conclusion that all or part of the required sequences are contained in the H-J segment. The

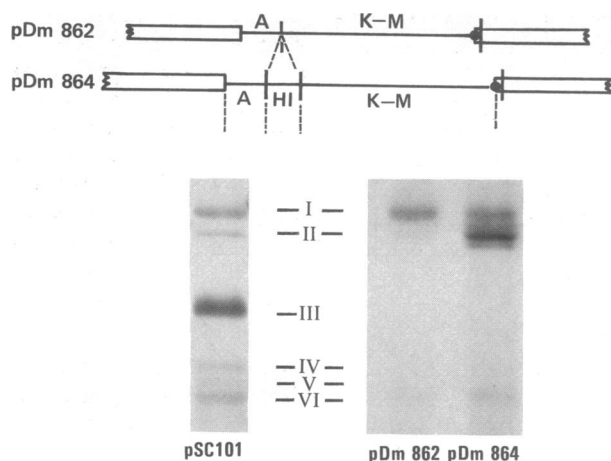


FIG. 4. Polypeptide synthesis promoted by the H-I segment. pDm864 was constructed, cloned, and mapped as described for pDm862 in Fig. 2, and the symbols used in that figure apply here. The orientation of the H-I insertion in pDm864 is not known.

following two experiments strongly suggest that the coding sequences for the Dm107 protein are confined to the H-J segment.

The ligation of the *Hind*III digestion products of pDm107 produces complex as well as simple deletions. pDm864 is such a complex deletion, and Fig. 4 shows that it differs from the simple pDm862 deletion by the insertion of the H-I segment, yielding a double deletion lacking the B-G and J segments. A comparison of the autoradiographs of the labeled polypeptides produced by pDm862 and 864 in AR1062 minicells indicates that the H-I insertion is responsible for the formation of two new polypeptide bands that migrate just in front of the pSC101 bands I and II—one strong band with an apparent molecular weight of 33,000 and a weaker band at 32,000.

The simplest explanation for these new polypeptides is that they represent truncated Dm107 proteins whose translation is initiated normally in the H segment, proceeds through the I segment, and is prematurely arrested because the J segment is absent. If it is assumed that "false" translation of sequences lying outside of the H-I insertion contributes little to the 33,000-dalton polypeptide, then approximately 45 amino acids at the carboxyl end of the Dm107 protein should be coded by sequences in J; and, because only 135 of the approximately 345 codons required for 38,000-dalton Dm107 protein can be contained in I, the remaining 165 codons should be in H. Any contribution of false translation to the 33,000-dalton polypeptide will alter the calculation to increase the codons in J and decrease them in H. For example, the median expectation for terminator codons that are randomly distributed among the falsely translated sequences is that they will cause termination prior to the false translation of 15 codons, and application of this expectation will change the calculated number of codons in H and J to 150 and 60, respectively.

This explanation leads to a prediction that has been tested regarding the polypeptides expected to be produced by the A-D, E-H, and I-M *Eco*RI fragments of Dm107 when each is inserted into pSC101 (Fig. 5). Clearly, only the E-H plasmid should yield a new polypeptide, and its molecular weight should be approximately 18,000, the exact value depending upon the relative amounts of false translation for this and the pDm864 plasmid. Fig. 5 shows that each of four independently cloned hybrids containing the E-H fragment yields two bands migrating in advance of the pSC101 band VI—an intense band at 17,000 daltons and a weaker one at 18,000. It is curious that

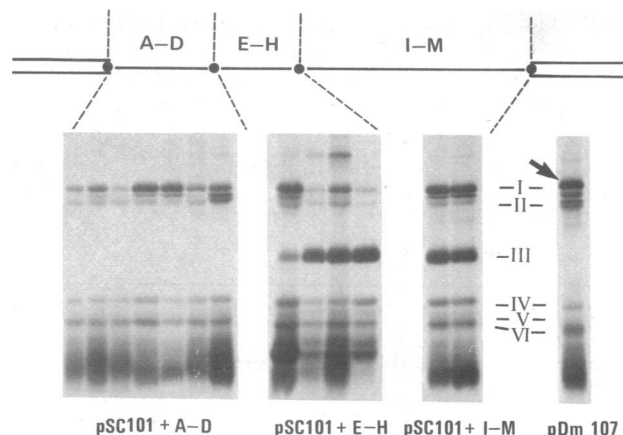


FIG. 5. Polypeptide synthesis promoted by the three *Eco*RI fragments of Dm107. The A-D, E-H, and I-M *Eco*RI fragments shown on the map of pDm107 at the top of the figure were cloned in pSC101 by ligating a complete *Eco*RI digest of pDm107 (100 μ g/ml) and using the products to transform HB101 to Tc^R as described in Fig. 2. The individual autoradiographs were obtained from independently cloned hybrids that had each been transferred to AR1062. The orientation of the *Eco*RI fragment in the different hybrids has not been determined.

a pair of polypeptides differing in molecular weight by 1000 is produced by both this and the preceding presumptive truncation of the Dm107 protein. Perhaps the shorter member of each pair derives from the longer by loss of a common segment that is made sensitive to proteolytic cleavage by incomplete folding of the truncated polypeptides.

It should be emphasized that no other new polypeptide was produced by the hybrids containing the *Eco*RI fragments. The band that is seen between pSC101 bands I and II in all the autoradiographs of Fig. 4 does not result from the plasmids because it is also seen with the AR1062 control lacking plasmids, when, as was the case here, the second sucrose gradient centrifugation is omitted during purification of the minicells (see Fig. 1 legend). The band seen near the top of all the autoradiographs can be similarly discounted.

DISCUSSION

The linkage between the vector and Dm segments in the hybrid plasmids introduces some uncertainty in assigning the coding sequences for new polypeptides to the Dm segment. A new polypeptide may derive from translation of codons on both sides of a joint linking the two segments, or the joints may themselves create new initiation signals that lead to the translation of codons in either segment. We have eliminated these possibilities with respect to the synthesis of the Dm107 protein by excising the internal C-K fragment from the Dm107 segment, inserting this fragment at a different vector site in each of the two orientations, and then demonstrating that the two resulting hybrids retain the capacity to synthesize this protein (Fig. 3).

This result also restricts the *D. melanogaster* DNA sequences required for the synthesis of the Dm107 protein to the C-K fragment. These sequences have been further localized to the C and H-K regions by the observation that the capacity for this synthesis is not lost by deletion of the D-G fragment from pDm107 (Fig. 2). By contrast, deletion of the H-J region results in the loss of this capacity. Hence we know that the H-J region contains required sequences, whereas the C and K segments may, but do not necessarily, contain such sequences.

When the I and I-J segments are deleted from the right end of the critical H-J region, progressively shorter polypeptides

of 33,000 and 17,000 daltons are synthesized in place of the 38,000-dalton Dm107 protein (Figs. 4 and 5). This shortening of the polypeptides is nicely correlated with the deletion lengths and, when taken in conjunction with the other mapping data, provides a strong argument that the Dm107 protein is translated from sequences within the H-J region—starting near the center of the 0.8-kb H segment, proceeding through the 0.4 kb of sequences in I, and finishing 0.1 to 0.2 kb within J.

Transcription of these coding sequences must then proceed from a promoter located to their left. That such a promoter resides within the C-K fragment is suggested by the equivalent amounts of the Dm107 protein generated by this fragment in both orientations (Fig. 3). The direction of transcription and the observation that deletion of the D-G region (Fig. 2) and of the B-G region (Fig. 4) does not prevent it, would further localize this promoter to the left half of the H segment, provided that the effective promoter in all of these hybrids is the same. Because we do not know that this is the case, such a placement is provisional.

Are these coding sequences transcribed and translated in *D. melanogaster* cells? We do not know. However, we think it unlikely that such a long row of approximately 345 sense codons would be maintained in the *D. melanogaster* genome unless it were expressed. We do not wish to imply that the promoter, the primary transcript, the mRNA, or indeed the final protein product will be the same as those in *E. coli*. Rather we think that the coding sequence in the H-J region and the amino acid sequence in the Dm107 protein will overlap those in the RNAs and polypeptides produced by some *D. melanogaster* cells, and hence, that the kind of experiments contained in this paper represent a potential route for identifying certain genes contained in cloned eukaryotic DNAs.

Our data allow us to map another kind of sequence in Dm107—namely, that responsible for the diminution of the synthesis of the pSC101 band III polypeptide, a diminution that has been observed for many of the pDm hybrids, including pDm107 (Fig. 1). Fig. 5 shows that such sequences reside in the A-D region, and examination of other pDm107 deletions (e.g., pDm862; Fig. 4) indicates they can be further localized to the A segment. The fact that the sequence-specific effect of the Dm insertions on the synthesis of this polypeptide is quantitative rather than qualitative (Fig. 1) suggests that it results from a change in the concentration of transcripts containing its coding sequences, not from an interruption of their translation by an insertion among them. Sequences in the Dm segments could effect this change by altering the properties of a pSC101 promoter at or near the *Eco*RI site of insertion, or by providing signals either for the initiation of transcription of the band III gene in the wrong direction or for the termination of transcription that must proceed through the insert to transcribe this gene in the right direction.

Finally, we note that the four new polypeptides that we detected after testing 37 pDm hybrids that contain 400 kb of *D. melanogaster* DNA require approximately 3.8 kb of coding sequences for their translation. Evidently the fraction of *D. melanogaster* sequences that can be efficiently translated in *E. coli* K-12 is small. Even our estimate of 1×10^{-2} for this proportion may be too large if the coding sequences for the

polypeptides generated by pDm102, 820, and 829 do not reside in the respective Dm segments. Others have recently observed new polypeptides in minicells containing pDm hybrids, but were unable to show that their codons reside in the Dm segment (18, 19); indeed, Meagher *et al.* (19) suggest that these codons derive, at least in part, from the vector DNA. While this estimate involves other uncertainties, including the fact that our scoring criteria will exclude certain polypeptides, we think it represents a useful first approximation, particularly because the other methods that have been used to detect the expression of eukaryotic DNAs in *E. coli* K-12 require extreme selective pressures (21, 22).

This work was supported by research grants from the National Science Foundation and from the National Institutes of Health. A.R. was both an Eleanor Roosevelt Fellow of the American Cancer Society and a Jane Coffin Childs Fellow.

1. Wensink, P., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) *Cell* 3, 315–325.
2. Glover, D. M., White, R. L., Finnegan, D. J. & Hogness, D. S. (1975) *Cell* 5, 149–155.
3. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961–3965.
4. Rubin, G., Finnegan, D. J. & Hogness, D. S. (1976) in *Progress in Nucleic Acid Research and Molecular Biology*, ed. Cohn, W. E. (Academic Press, New York), Vol. 19, pp. 221–226.
5. Glover, D. M. & Hogness, D. S. (1977) *Cell* 10, 167–176.
6. White, R. L. & Hogness, D. S. (1977) *Cell* 10, 177–192.
7. Young, M. W. & Hogness, D. S. (1977) in *ICN-UCLA Symposia on Molecular and Cellular Biology*, eds. Wilcox, G., Abelson, J. & Fox, C. F. (Academic Press, New York), Vol. 6, in press.
8. Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* 41, 459–472.
9. Adler, H. I., Fisher, W. D., Cohen, A. & Hardigree, A. A. (1966) *Proc. Natl. Acad. Sci. USA* 57, 321–326.
10. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. (1973) *Proc. Natl. Acad. Sci. USA* 70, 3240–3244.
11. Glover, D. M. (1976) in *New Techniques in Biophysics and Cell Biology*, eds. Pain, R. & Smith, B. (John Wiley, London), Vol. 3, pp. 125–145.
12. Greene, P. J., Betlach, M. C., Goodman, H. M. & Boyer, H. W. (1974) *Methods Mol. Biol.* 7, 87–111.
13. Wilson, G. A. & Young, F. E. (1975) *J. Mol. Biol.* 97, 123–125.
14. Smith, H. O. & Wilcox, K. W. (1970) *J. Mol. Biol.* 51, 379–391.
15. Roozen, K. J., Fenwick, R. G., Jr. & Curtiss, R., III (1971) *J. Bacteriol.* 107, 21–33.
16. Adams, M. H. (1959) *Bacteriophages* (Interscience Publishers, New York).
17. Chang, A. C. Y., Lausman, R. A., Clayton, D. A. & Cohen, S. N. (1975) *Cell* 6, 231–244.
18. Miller, D. L., Gubbins, E. J., Pegg, E. W., III & Donelson, J. E. (1977) *Biochemistry* 16, 1031–1038.
19. Meagher, R. B., Tait, R. C., Betlach, M. & Boyer, H. W. (1977) *Cell* 10, 521–536.
20. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. (1974) *Proc. Natl. Acad. Sci. USA* 71, 3455–3459.
21. Struhl, K., Cameron, J. R. & Davis, R. W. (1976) *Proc. Natl. Acad. Sci. USA* 73, 1471–1475.
22. Ratzkin, B. & Carbon, J. (1977) *Proc. Natl. Acad. Sci. USA* 74, 487–491.