A new gene nested within the *dunce* genetic unit of *Drosophila melanogaster*

Maria Furia[1,2]*, Filomena A.Digilio[2], Dora Artiaco[2], Ennio Giordano[1] and Lino C.Polito[1,2]
[1]Dipartimento di Genetica, Biologia Generale e Molecolare, Università di Napoli and [2]Istituto di Biochimica delle Proteine ed Enzimologia, CNR, Naples, Italy

Received April 12, 1990; Revised and Accepted June 28, 1990 EMBL accession no. X52118

ABSTRACT

The molecular organization of the *dunce* gene of *Drosophila melanogaster* has proved to be particularly complex, with two divergently transcribed genes, *Sgs-4* and *Pig-1*, nested within its 79 kb intron (1). Here we report the identification and the molecular characterization of a third gene nested within the transcription unit of *dunce*. This newly identified gene is located nearly 6 kb downstream *Pig-1*, within a more upstream *dunce* intron. The gene is developmentally regulated and transcribed with the same polarity of *dunce*; several lines of evidence indicate that it might encode for a salivary gland secreted (Sgs) protein.

INTRODUCTION

The *dunce* gene of *D. melanogaster* plays an important role in the learning and memory processes of the fly (2–7). *dunce* mutants display an abbreviated memory and a variety of behavioural defects, correlated with a deficiency in the activity of one cAMP phosphodiesterase form (8). The gene maps on the X chromosome; by using mapped chromosomal breakpoints, it has been located at chromomere 3D4, where exons 3–13 were subsequently assigned (8, 9). However, the isolation of a *dunce* cDNA clone, named 863, has recently revealed that two additional exons, designated 1 and 2, reside at chromomeres 3C11–12 (1). Although the molecular cloning of the gene has so far not been completely achieved, its organization is particularly elaborate. Two genes, *Sgs-4* and *Pig-1*, are in fact located within the 79 kb intron that separates exon 3 from exon 2.

In this paper we report the identification of another gene nested within *dunce*, a finding that adds further complexity to the transcriptional pattern of the region. The newly identified gene is intronless and encodes a poly(A)+ mRNA of about 0.6 kb which is abundant only during the third larval stage and is salivary gland specific. The protein encoded by the nested gene shares many features with the salivary gland secreted (Sgs) proteins described so far, suggesting that the gene might be a new member of the well characterized Drosophila *Sgs* genes family.

MATERIALS AND METHODS

Northern analyses

For Northern analyses, the flies were grown at 21°C and total RNA was extracted at different developmental times as described by Pirrotta *et al.* (10). The RNA was fractionated on 1.2% agarose formaldehyde gel, transferred to a nylon membrane (Hybond N, Amersham) according to Thomas (11), hybridized and washed following the Amersham suggested protocols.

DNA probes labelling

For both Northern analyses and library screening, the DNA probes consisted of DNA fragments isolated from low melting temperature agarose and labelled to a specific activity of at least 5×10^8 cpm/μg by the Multiprime labelling kit (Amersham).

cDNA library screening and DNA sequencing

A cDNA library prepared in the lambda gt10 vector from *D. melanogaster* third instar larvae cytoplasmic poly(A)+ RNA was kindly provided by S.Schneuwly. The library was screened with the 3 kb HindIII genomic fragment described in the text and derived from a genomic library previously obtained from an Oregon R strain (12).

For the DNA sequencing, cDNA and genomic fragments were cloned in pUC18 plasmid and nucleotide sequences determined by the dideoxy method (13). Sequencing data were analyzed by the Micro Genie sequence analysis program.

Nuclease S1 mapping

For the nuclease S1 protection assay, the double-stranded 2 kb XhoI fragment described in the text was terminally labelled at the 5′ end with (gamma-^{32}P) ATP and T4 polynucleotide kinase; 1×10^5 cpm of the labelled fragment were hybridized to 200 μg of total third larval instar RNA in 80% formamide solution at 49°C for 4 hrs. RNA/DNA hybridization conditions and nuclease S1 digestion were essentially as described by Maniatis *et al.* (13). S1 protected fragments were analyzed on 6% polyacrylamide gels containing 8M urea.

* To whom correspondence should be addressed

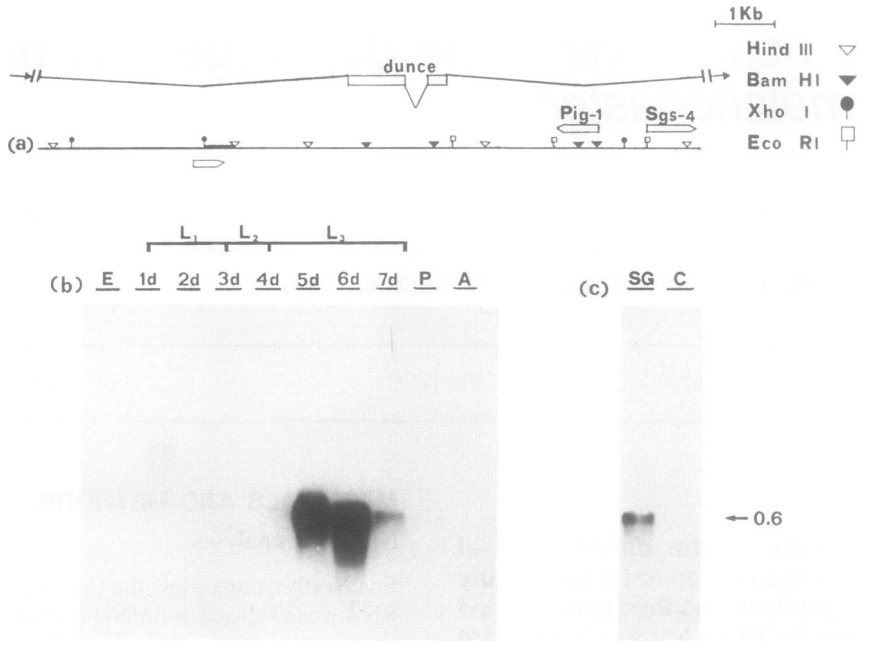

Fig. 1. Genomic localization of the 0.6 kb transcript (a) and its developmental and tissue expression specificity (b), (c). (a) Position of the 0.6 kb transcription unit aligned on a genomic partial restriction map. On the top, relative positions of *dunce, Sgs-4* and *Pig-1* genes. (b) A Northern developmental blot in which 40 μg of Oregon R total RNA extracted at different times of development were loaded in each lane. E, 8–10 hr embryos; 1d–7d, RNA preparations obtained every 24 hours following embryo collection (L1, first instar larvae; L2, second instar larvae; L3, third instar larvae; 1d, 3d and 4d collections may contain populations at two developmental stages; P, pupae; A, adults). The probe used in the experiment is the 0.5 kb HindIII/XhoI genomic fragment indicated on the map by the thick line. (c) Tissue specificity of the 0.6 kb transcription unit. 20 μg of total RNA extracted from manually dissected salivary glands (SG) and from the remaining larval carcasses (C) were hybridized to the same probe described in (b).

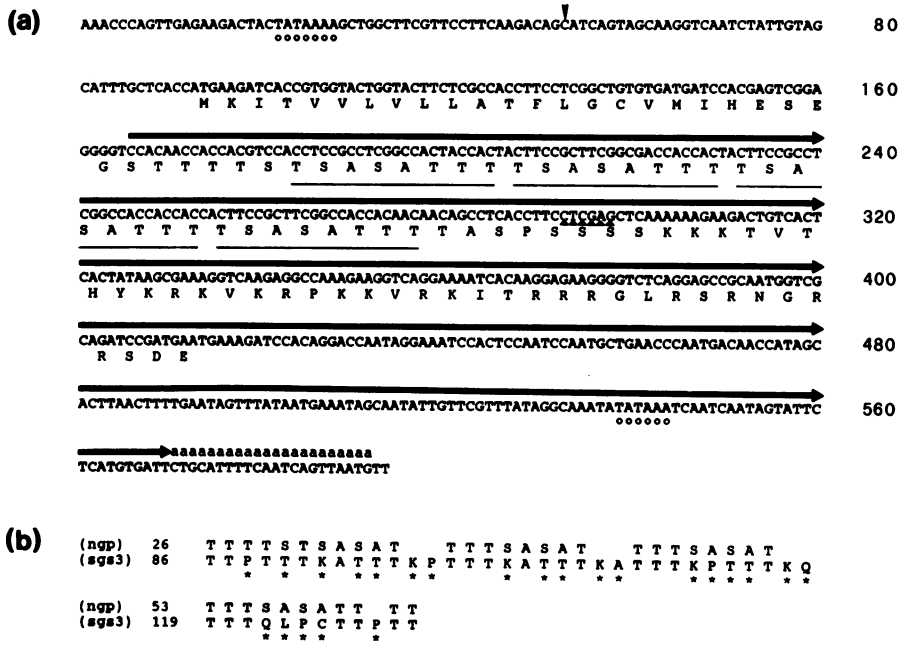

Fig. 2. Sequence analysis of the SC1 cDNA clone. (a) A comparison of the sequence of the 3 kb HindIII genomic fragment with the SC1 sequence in the region of the overlap. The extent of SC1 is shown by the thick line above the genomic sequence, while the orientation of the transcription is indicated by the horizontal arrows and the position of the poly(A) track. The postulated TATA box and the poly(A) addition signal are underlined by open circles. The vertical arrow marks the main transcription start site determined on the basis of the S1 nuclease assay shown in Fig. 3; the position of the protected XhoI site is underlined by black triangles. The predicted protein sequence is reported and the 8-amino acid repeats are underlined. (b) The nested gene protein sequence (ngp) is compared to the Sgs-3 protein, with an alignment showing the 48.9% of homology. The numbers on the left indicate the positions of the amino acid residues and asterisks on the bottom mark the differences.

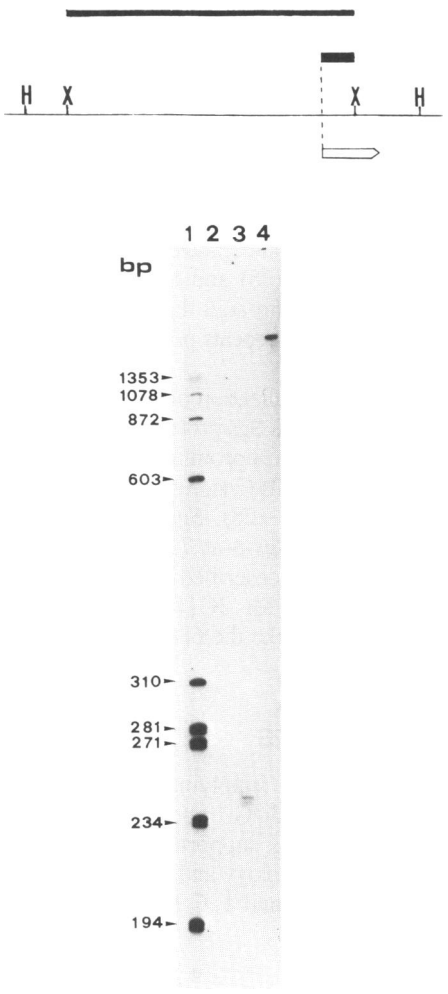

Fig. 3. Localization of the 5′ end of the 0.6 kb mRNA by S1 mapping. On the top, diagrammatic representation of the 2 kb XhoI genomic fragment used in the S1 mapping experiment and the overlapping protected fragment. The 5′ end labelled XhoI fragment was hybridized to third larval instar total RNA; after S1 nuclease treatment the protected fragments were loaded on 6% polyacrylamide-urea gel (lane 3). The S1 untreated XhoI fragment (lane 4) and the same fragment treated with S1 in absence of RNA (lane 2) were loaded as control. The molecular weight marker was HaeIII digested φX174 (lane 1). The 245−240 nt long protected fragments obtained have the sizes expected for an RNA molecule starting 25−30 nt downstream of the genomic promoter sequence indicated in Fig. 2.

RESULTS

Identification of the gene

In the course of a study aimed at analyzing the transcriptional activity of the *D. melanogaster* 3C11−12 intermolt puff region, we identified by Northern blot experiments an RNA molecule, nearly 0.6 kb in size, which originates within the puff region, about 6 kb downstream of the 3′ end of *Pig-1* mRNA. The transcript was detected in the total and poly(A)+ third larval instar RNA by hybridization with the 3 kb HindIII genomic fragment shown in the map (Fig. 1a). The fragment is placed on the left of *dunce* exon 1, within an upstream *dunce* intron (1). To check whether the identified transcript was developmentally regulated, total RNA was extracted at each stage of the Drosophila life cycle, from embryos to adults, and RNA blots hybridized to a probe derived from the HindIII fragment described above. The temporal Northern profile thus obtained clearly shows that the transcript is detected only during the third larval instar, reaching its maximum level at the onset and lasting till the end of this stage (Fig. 1b).

It is noteworthy that, as in the case of the mRNAs encoded by the close *Sgs-4* and *Pig-1* genes, the 0.6 kb RNA molecule is detected only in the salivary glands (Fig. 1c).

cDNA clone isolation and gene organization

To look at the genomic organization of the 0.6 kb transcript we screened a third larval instar cDNA library, using as probe the 3 kb HindIII genomic fragment described above. Three cDNA clones representative of the 0.6 kb RNA molecule were recovered and analyzed by restriction mapping, hybridization on genomic and RNA blots and nucleotide sequence analysis. The sequence of the longest cDNA clone obtained, named SC1, is reported in Figure 2, aligned with the corresponding genomic sequence. SC1 is 427 bp long and is perfectly colinear with its genomic region. The 3′ end of the 0.6 kb RNA appears to be fully represented, since the clone contains a 22 nt long poly(A) tail and a consensus polyadenylation sequence 26 nt upstream. The position of the poly(A) track establishes for the 0.6 kb RNA molecule the same transcription polarity as that of *dunce*, from the telomere towards the centromere.

The SC1 size, together with the sequence analysis of the genomic region, suggested that the clone did not contain the sequences representing the 5′ end of the transcript. A continuous open reading frame (ORF), 107 amino acids long, is in fact present on the genomic sequence, starting 73 bp upstream the SC1 5′ end (Fig. 2). Computer analysis revealed that this ORF showed the codon bias typical of Drosophila genes (14); in addition, the sequence upstream the first methionine codon matched satisfactorily (75%) the proposed consensus sequence flanking the translation starting sites in *D. melanogaster* (15). These findings supported the hypothesis that the ORF was actually translated and, therefore, was likely to be entirely contained within the 0.6 kb transcript.

In order to define the 5′ end of the 0.6 kb RNA molecule, an S1 mapping experiment was performed. When looking at the upstream genomic sequence, we noticed that a canonical TATA box was present at position −64 from the first methionine codon of the ORF discussed above that could direct the transcription initiation. The 2 kb XhoI genomic fragment containing the putative transcription start site was then terminally labelled and used in the S1 nuclease protection assay. Two closely spaced fragments, about 245−240 nt long, were protected (Fig. 3), revealing some heterogeneity in the transcription initiation. However, the S1 mapping essentially confirms the start site expected on the basis of the genomic sequence (Fig. 2) and clearly establishes that the entire ORF is actually contained within the 0.6 kb RNA.

Taken together with the 3′ end assignment, the S1 mapping of the 5′ end indicates that the mRNA extends continuously for 513−518 nt, plus the length of the poly(A) tail. This length is in good agreement with the mRNA size previously estimated from RNA blots, and predicts a poly(A) tail of about 80 nt.

Structure of the nested gene protein

The protein encoded by the 0.6 kb mRNA is 107 amino acids long, with a predicted molecular weight of 11.3 kd.

The protein is unusually rich in threonine and serine (Thr +

Ser = 41%), both amino acids representing potential sites of glycosylation. 15 hydrophobic residues are present within the first 20 amino terminal positions of the predicted protein sequence, suggesting the occurrence of a putative signal peptide. The amino terminal region of the protein is particularly rich in threonine and is internally redundant, showing the 8-amino acid units Thr-Ser-Ala-Ser-Ala-Thr-Thr-Thr tandemly repeated 4 fold, from residue 31 to 62 (Fig. 2). Interestingly, a computer search showed that this segment was highly homologous (48.9%) to the internal redundant region of the Sgs-3 protein of *D. melanogaster* (16), whose gene is located on chromosome 3, at the intermolt puff 68C3−4 (Fig. 2b).

It has been reported that the Sgs3 protein shows a modular organization, with a first module composed by the leader signal peptide, an internal module highly rich in threonine, and a cysteine-rich module towards the carboxy-terminal end. The Sgs-3 internal module is in turn composed of an amino-terminal threonine-rich region, followed by an internally redundant segment, still threonine-rich, containing imperfect tandem repeats of a 5-amino acid basic unit Pro-Thr-Thr-Thr-Lys. It has been suggested that a possible function of this internal threonine-rich module is to furnish targets for the extensive glycosylation shown by the Sgs-3 protein (16).

Although lacking the carboxy-terminal cysteine-rich module, the organization of the nested gene protein (ngp) described here shares many features with the one exhibited by the Sgs-3 protein. It presents a signal peptide followed by a repeated internal module highly rich in threonine (50%), which shows both sequence and structure homology with the Sgs-3 internal module.

75% of the residues within the ngp 8-amino acid repeats are potentially glycosylatable sites (serine and threonine), suggesting that the ngp redundant region may play a role similar to that proposed for the Sgs-3 internal module.

DISCUSSION

The experiments reported in this paper demonstrate the existence of another Drosophila gene with a nested arrangement. To verify that the newly identified gene is actually nested within the *dunce* transcription unit we pursued two different approaches. First, we checked whether the SC1 complementary clone consisted of single copy sequences. Second, we controlled that the gene was actively expressed in different Drosophila stocks, including Canton S, the only strain in which the position of the *dunce* exons 1 and 2 has so far been established (1). Both approaches confirmed the intronic localization of the 0.6 kb transcription unit (data not shown).

Hence, in addition to *Sgs-4* and *Pig-1* genes, at least another genetic unit is nested in a *dunce* intron.

The number of eukaryotic genes with a complex and overlapping organization is rapidly increasing. Since the first reported example of nested genes at the *Gart* locus in Drosophila (17), overlapping transcripts have been described for a mouse locus (18), for the *Ddc* (19) and *Tpr* (20) loci of Drosophila, for the *GnRH* locus of the rat (21) and, more recently, for the *herbA* homologue locus in the human genome (22). The organization of the *dunce* locus in *D. melanogaster* so far appears to be the most complex, with at least three genetic units lying in two different introns, the *Sgs-4* and the nested gene described here encoded by the same strand of *dunce* and *Pig-1* by the opposite.

Several lines of evidence indicate that the gene identified in our experiments encodes for a salivary gland secreted protein.

Firstly, the gene shows the same developmental and tissue specific expression profile displayed by the other well characterized Drosophila *Sgs* genes. Secondly, the predicted protein sequence shares many features with the Sgs proteins described so far. The protein in fact presents a putative signal peptide, as expected for a secreted protein, contains many potentially glycosylatable sites and is internally redundant, as described for both Sgs-3 (16) and Sgs-4 (23) proteins.

Furthermore, the 8-amino acid threonine-rich repeat is very similar to the 5-amino acid repeats present in the internal module of the Sgs-3 protein.

Lastly, the genomic localization of the gene also supports the hypothesis that it encodes an Sgs protein. In fact, all the *Sgs* genes have been mapped at the most prominent intermolt puff loci, with the *Sgs-1* gene located at 25B (24), *Sgs-3, Sgs-7* and *Sgs-8* located in cluster at 68C3−4 (25−28), *Sgs-4* at 3C11−12 (29−30), *Sgs-5* at 90BC (31) and *Sgs-6* at 71C3−4 (32).

The putative *Sgs* gene described here is located within the genomic region required for 3C11−12 puff formation (33), suggesting that this puff, like the 68C3−4, might contain more than one *Sgs* gene.

ACKNOWLEDGEMENTS

We thank S.Schneuwly for supplying us with an aliquot of third instar cDNA library and G.Imperato for skilful technical assistance. This work was supported by a grant of the CNR Target Project on Biotechnology and Bioinstrumentation to M.F. and MPI grants to M.F. and L.C.P.

REFERENCES

1. Chen,C., Malone,T., Beckendorf,S.K. and Davis,R.L. (1987) *Nature* **329**, 721−724.
2. Aceves-Pina,E.O., Booker,R., Duerr,J.S., Livingstone,M.S., Quinn,W.G., Smith,R.F., Sziber,P.P., Tempel,B.L. and Tulley,T.P. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **48**, 831−840.
3. Duday,Y. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5445−5448.
4. Tempel,B.L., Bonini,N., Dawson,D. and Quinn,W.G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1482−1486.
5. Tulley,T. and Quinn,W.G. (1985) *J. Comp. Physiol.* **157**, 263−277.
6. Kyriacou,C.T. and Hall,J.C. (1984) *Nature* **308**, 62−64.
7. Dudai,Y., Jan,K., Byers,D., Quinn,W.G. and Benzer,S. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1684−1688.
8. Chen,C.N., Denome,S. and Davis,R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9313−9317.
9. Salz,H.K., Davis,R.L. and Kiger,J.A.Jr. (1982) *Genetics* **100**, 587−596.
10. Pirrotta,V., Hadfield,C. and Pretorius,G.H.J. (1983) *EMBO J.* **2**, 927−934.
11. Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201−5205.
12. Kaiser,K., Furia,M. and Glover,D.M. (1986) *J. Mol. Biol.* **187**, 529−536.
13. Maniatis,T., Fritsch,F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
14. O'Connell,P. and Rosbash,M. (1984) *Nucl. Acids Res.* **12**, 5495−5513.
15. Cavener,D.R. (1987) *Nucl. Acids Res.* **15**, 1353−1361.
16. Garfinkel,M.D., Pruitt,R.E. and Meyerowitz,E.M. (1983) *J. Mol. Biol.* **168**, 765−789.
17. Henikoff,S., Keene,M.A., Fechtel,K. and Fristrom,J.W. (1986) *Cell*, **44**, 33−42.
18. Williams,T. and Fried,M. (1986) *Nature* **322**, 275−279.
19. Spencer,C.A., Gietz,R.D. and Hodgetts,R.B. (1986) *Nature* **322**, 279−281.
20. Wong,F., Yuh,Z., Schaefer,E.L., Roop,B.C. and Ally,A.H. (1987) Somatic Cell and Mol. Genetics **13**, 661−670.
21. Adelman,J.P., Bond,C.T., Douglass,J. and Herbert,E. (1987) *Science* **235**, 1514−1517.

22. Miyajima,N., Horiuchi,R., Shibuya,Y., Fukushige,S., Matsubara,K., Toyoshima,K. and Yamamoto,T. (1989) *Cell*, **57**, 31–39.
23. Muskavitch,M.A.T. and Hogness,D.S. (1982) *Cell* **29**, 1041–1051.
24. Velissariou,V. and Ashburner,M. (1980) *Chromosoma* **77**, 13–27.
25. Korge,G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4550–4554.
26. Korge,G. (1977) *Chromosoma* **62**, 155–174.
27. Akam,M.E., Roberts,D.B., Richards,G.P. and Ashburner,M. (1978) *Cell* **13**, 215–225.
28. Crowley,T.E., Bond,M.W. and Meyerowitz,E.M. (1983) *Mol. Cell. Biol.* **179**, 289–314.
29. Korge,G. (1977) *Develop. Biol.* **58**, 339–355.
30. Muskavitch,M.A.T. and Hogness,D.S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7362–7366.
31. Guild,G.M. and Shore,E.M. (1984) *J. Mol. Biol.* **179**, 289–314.
32. Velissariou,V. and Ashburner,M. (1981) *Chromosoma* **84**, 173–185.
33. McNabb,S.L. and Beckendorf,S.K. (1986) *EMBO J.* **5**, 2331–2340.