A protein product of the *Drosophila* recessive tumor gene, *l* (2) *giant gl*, potentially has cell adhesion properties

Reinhard Lützelshwab, Christian Klämbt, Rita Rossa and Otto Schmidt

Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg, FRG

Communicated by H.Kössel

In the fruitfly *Drosophila melanogaster* the recessive mutation *lethal* (2) *giant larvae* affects cell growth and cell adhesion properties of neural and imaginal disc cells during embryonic and larval development. Here we report the nucleotide sequence of a cDNA from one of the two major transcripts of the *l* (2) *gl* gene containing a 3.5-kb open reading frame. From the deduced amino acid sequence we predict a 130-kd protein with certain sequence similarities to proteins with cell adhesion properties. We have raised antibodies against bacterial fusion proteins encoded by cDNA subfragments. Affinity-purified specific antibodies recognize a 130-kd protein in wild-type protein extracts and *in situ*, which is missing in the mutant. Localization of the *l* (2) *gl* protein in tissue sections indicates that the *l* (2) *gl* protein is associated with cellular membranes and probably involved in cell–cell interaction.

Key words: *Drosophila*/recessive tumor gene/cDNA/antibodies/cell adhesion

Introduction

During *Drosophila* early development several gene systems participate in molecular decision processes that define the state of determined cells (Garcia-Bellido, 1975; Lewis, 1978; Nüsslein-Volhard and Wieschaus, 1980; Gehring, 1985; Mahowald and Hardy, 1985). Cell growth and tissue specificity are evident manifestations of these ongoing determination processes that can already be detected in the differential proliferation rates and cell adhesion properties of imaginal disc cells (Garcia-Bellido, 1966; Postlethwait and Schneiderman, 1971; Gauger *et al.*, 1985). *Drosophila* larvae with the recessive mutation *lethal* (2) *giant larvae* (*l* (2) *gl*) show unstructured growth of the brain and imaginal discs (Gateff and Schneiderman, 1967). When subcultured in the abdomen of adult females, homozygous mutant cells grow faster and in a more dispersed fashion compared with the wild-type and eventually kill the host flies (Gateff and Schneiderman, 1969, 1974; Gateff, 1978b). The phenotypic effects are retained in subcultured mutant cells, their growth characteristics being comparable with tumors of vertebrates (Gateff, 1978a,b).

The *l* (2) *gl* gene was recently cloned (Mechler *et al.*, 1985). Two transcripts, one of 4.3 kb and another of 5.5–6.0 kb in length, were identified on Northern blots. Both are found to be expressed biphasically in early embryogenesis and again in late third instar larvae (Mechler *et al.*, 1985). Both transcripts were absent in larval tissue of mutant strains (Lützelshwab *et al.*, 1986).

To analyse the molecular function of the wild-type gene products of the *l* (2) *gl* gene locus we have isolated cDNA fragments

and characterized the proteins from the gene. Here we report the nucleotide sequence of an almost complete cDNA of the large transcript including an open reading frame of 3479 bp. The deduced amino acid sequence revealed structural features which are similar to proteins involved in cellular interaction. Together with results from the analysis of antibodies raised against bacterial fusion proteins we conclude that this *l* (2) *gl* protein is located at the cellular membrane or in the intercellular matrix and is potentially involved in cellular interaction of determined cells.

Results

Transcripts and cDNA

Cloned genomic DNA fragments from the transcribed region of the *l* (2) *gl* gene (Mechler *et al.*, 1985; Lützelshwab *et al.*, 1986) were used to screen a phage library of cloned *Drosophila* cDNA fragments (Poole *et al.*, 1985). To screen for large cDNA fragments we used genomic subfragments from the inferred 5' half of the transcribed region of the gene (Lützelshwab *et al.*, 1986). Positive isolates were subcloned into the plasmid vector phs-neo (Steller and Pirrotta, 1984) and its derivatives were analysed. The cDNA clones pE7-9, pE7-19 and pE7-20 were from separate recombinant phages containing similar-sized inserts of ~5.3 kb length (Figure 1 for pE7-9). Comparison of restriction sites of cDNA clones with the corresponding genomic DNA showed that the primary transcript is spliced several times to give the mature mRNA.

To deduce the nucleotide sequence of the cDNA fragment, we analysed the DNA sequence of the cDNA clone pE7-9 and of the genomic DNA from the 3' half of the gene, where almost all of the coding sequences are located. Comparison of cDNA and genomic sequences so far analysed indicate that the coding sequences in the 3' half of the gene are interrupted by five intervening sequences from 65 to 500 bp in length.

The nucleotide sequence of the cDNA pE7-9 revealed an open reading frame of 3479 bp (Figure 2). A 5'-distal sequence of 396 bp contains stop codons in all three possible reading frames (Figure 2). The two possible initiation codons, one AUG(U) codon at position 396 and an additional AUG(G) at position 540, are both in the open reading frame (Figure 2). If the initiation of translation is assumed at position 540, which resembles the eukaryotic translation initiation consensus sequence (Kozak, 1984), a protein of 1114 amino acids in length is predicted for the translation product of this transcript. This corresponds to an expected mol. wt of ~130 kd.

To verify the open reading frame found by nucleotide sequence analysis we have tested these sequences in bacterial expression vectors. cDNA subfragments (Figure 1) were subcloned into the expression vector pATH (see Materials and methods). The mol. wt of the fusion protein obtained from each of the respective recombinant expression plasmids was determined and shown to correspond to the coding capacity of the cloned subfragment (C. Klämbt, unpublished results). This analysis, which covered most of the open reading frame sequences, confirmed the existence of the above-mentioned protein-coding sequence.

Fig. 1. Restriction map of the cDNA clone pE7-9. The 5.4-kb cDNA fragment is from the larger transcript seen on Northern blots (Mechler *et al.*, 1985; Lützelschwab *et al.*, 1986). Open reading frame sequences on pE7-9 are indicated by open boxes. Two AUG sequences which could act as possible initiation codons (see Figure 2) are indicated (arrows). pE7-9.A is a *Bal31* truncated fragment of pE7-9. pPTH6, pPTB5, pPTH2 and pPTBa1 are subfragments of the cDNA fragment cloned into the bacterial expression vector pATH (see Materials and methods). B, *Bam*HI; H, *Hind*III; R, *Eco*RI.

A transmembrane signal peptide at nucleotide position 679 (Figure 2) was found with a 13-amino acid sequence (IleLeuSerLeuThrAlaAlaAsnGlnLeulleuTrp) similar to the consensus sequence (Perlman and Halvorson, 1983) with a basic residue (Arg) preceding the hydrophobic portion of the signal sequence (Steiner *et al.*, 1980; Inouye and Halegoua, 1980). Although the signal peptide is not located at the immediate amino terminus of the predicted protein, a functional transmembrane signal at a more internal position (Spiess and Lodish, 1986) would predict that the protein is transported through the cellular membrane. The distribution of hydrophobic amino acids indicates no extended hydrophobic regions which could act as a transmembrane sequence (Figure 3). It is possible that the protein is fixed to the membrane by the transmembrane signal peptide (Spiess and Lodish, 1986) or that the protein is cleaved from the signal peptide and deposited into the extracellular space.

Analysis of the deduced protein sequence revealed internal repeats of similar amino acid sequences (Figure 4). The location of the repeated features within the hydropathy pattern and the spacing of cysteine residues indicate that some of the repeats consist of alternating stretches of predominantly hydrophilic then hydrophobic amino acid residues, bordered by cysteine residues in most cases (Figure 3). Within the repeats are regularly spaced potential glycosylation sites, charged amino acids and cysteine residues (Figure 4).

A large middle region shows sequence similarity among all the repeats, whereas the distal parts of some of the repeated sequences contain amino acid residues conserved within a subfamily of repeats, suggesting a closer evolutionary or functional relatedness (Figure 4).

Within repeat 4 a cell-attachment recognition signal RGDV, which might have a functional significance in *Drosophila* embryogenesis (Naidet *et al.*, 1987), is located between two cysteine residues (Figure 4).

One common feature among the repeats is that negatively charged amino acids and asparagine residues (some of them being potential *N*-glycosylation sites) appear to be regularly spaced with respect to cysteine residues (Figure 5). A consensus sequence from this region present in most of the *Drosophila* repeats was used to screen for sequence similarities in proteins compiled in the Dayhoff and Doolittle protein sequence libraries (Kyte and Doolittle, 1982; Orcutt *et al.*, 1982). From this screen additional proteins with a similar peptide sequence from several organisms were identified (see Materials and methods). From 12 proteins identified with this scan, nine are cell surface or viral proteins.

The corresponding sequences of three proteins which are assumed to be located on the cell surface, like the contact site A protein from *Dictyostelium* csA (Noegel *et al.*, 1986), the neural cell adhesion molecule N-CAM (Hemperly *et al.*, 1986) and integrin (Tamkun *et al.*, 1986), are included in Figure 5. In the deduced amino acid sequences of csA protein and N-CAM, additional structural similarities were found in the protein sequence adjacent to the consensus sequence used for the screen. Similarities of conserved features of the *Drosophila* repeats to the protein sequence of the *Dictyostelium* csA protein (Noegel *et al.*, 1986) led to our identification of repeated sequences within the csA protein (Figure 6). Some of the conserved amino acids within these repeats are spaced in a similar fashion in the *Drosophila* repeats (Figure 6) and in two of the N-CAM repeats (not shown). This suggests a structural relatedness of the three proteins which, together with the cellular location of the *Drosophila* protein (see below), could indicate that the 130-kD protein coded by the *l* (2) *gl* locus is potentially involved in cell adhesion.

Antibodies against *l* (2) *gl* protein

A 2.2-kb *Hind*III fragment from the 3' end of the cDNA fragment pE7-9 was cloned into the *Hind*III site of the bacterial expression vectors pATH10 (Körner, unpublished) and pUR291 (Rüther and Müller-Hill, 1983) to give the plasmids pTH6 and pPGH5, respectively. Fusion proteins were synthesized in transformed bacteria, and the mol. wts were estimated in SDS-polyacrylamide gels and compared with the size of the vector proteins. The size differences indicate that the protein coded by the *Drosophila* cDNA insert amounts to 30 kD in both fusion proteins (Figure 7a). This correlates with the expected length of the open reading frame on the 2.2-kb *Hind*III fragment from the cDNA (Figure 1). The fusion protein coded by the plasmid pPTH6 (*trpE* fusion protein) was used as an antigen to raise antibodies in rabbits (White and Wilcox, 1984). The antiserum obtained was tested on Western blots of electrophoretically separated bacterial extracts. It showed that both fusion proteins are labeled with the total serum (Figure 7b). In addition, other bacterial proteins are recognized by the antiserum. As expected, no cross-reaction of serum was observed with the β -galactosidase protein (Figure 7b). In order to purify antibodies directed against the *Drosophila* protein, a single purification step was carried out. In this procedure the alternative fusion protein encoded for by the pPGH5 plasmid (*lacZ* fusion protein) was isolated and covalently bound to a Sepharose 4B column. The total anti-*l* (2) *gl/trpE* fusion protein serum was applied, and after washing the bound antibodies were

1 CCCCCCGGAACACCTGAACCCACCGATCGGCACCTTTGTGTGTGTTAAATGACCGGCATATCGTTTGGGCTAATAATATTGTTACAGTTTATAAAACAGATACACGGTTAAGCTCTGA

121 CCGGTCGTATTATCATCTGTGTGTTCTTGGTGTGACACGTTGTTTTCGTTGTCTCATTAAAGCAACAGATGTGAAGAACAGTCGGGGGGTTCTCAACATAGAATACATACAAACTTA

241 TCGGTGTTGGTGAGCAATAAAATAGTCAAAAAATTACTCGGCCACCGCTTACACCGTCAAAACGGAAATACATAAAGTTTGAATAAAATACGGCCACAAATAGTATACATATTATTCT

361 ATCGAAACCAACCCACCACCAAGAGCATAACCAATTATGTTAAAGTTTATCAGAGGAAAGGGCAGCCGCTGTCAGACAGACCGCTCAGAGAAGCAAGCTTTGCTTATCGTAAGA
MetLeuLysPheIleArgLysGlnGlnProSerAlaArgHisGlyLeuGlnLysPheAlaIleArgLysThr

481 CGGCACGATGGCTTCCCTCATGCTTCGGCTTTCGGGTATGATCCAGCTTTTGAACCTTATGGCAATAGGGACGCAACAGGGGCTTTAAAGTTTTCGGTCAACCCGGAGTGAATTGT
AlaGlnHisGlyPhePheHisAlaSerAlaIleArgLysAlaIleArgLysLeuMetAlaIleGlyThrGlnThrGlyAlaLeuLysValGlyGlnProGlyValLeuThr

601 ACGGTCAGCATCTTTGTTAAACATTCAGCATCGGAGCTTAATGTACAACTTGTGAATGGGTGTGAAGCTGGGCTCACTTTCGTTGACGGCAGCGAATCAATTAATTCATGGG
GlyGlnHisThrLeuLeuAsnAsnSerAlaSerGluLeuAsnValGlnLeuLeuGluTrpValTyrGlyThrGlyArgGluLeuSerLeuThrAlaAlaAsnGlnLeuIleLeuTrpGlu

721 AGCCAGTTGGAGCAACGTTGCTGCCAATCAAAACACTACCGTTTTCAGCGCAACCTTAAAGGTTTTCATCGCTGCTGCTTCTCAGTAAGGATTCGTATGGATTGGAACAGAGGT
ProValGlyAlaIleLeuLeuProIleLysThrLeuProPheAspGlyLysLeuLysValSerLysSerCysSerLysLeuLysAspLeuThrProIleGlyThrGlyGly

841 GAAACCTTCTACATCGGATTATCATACATTTACCAATTAAGGCGCTGAATTTACCATGACGCTGTGCTAGACAGGCTGCCACAGGCTACAAGCTAAATCTAGTGTAAATGAGTCA
AsnIleThrLeuLeuLysPheHisThrPheThrIleLysGluProValIleIleThrHisGlyArgValValGluGlnValProAlaIleThrLysLysLeuAsnProSerAlaIleGlySerIle

961 TCCGCCAATCTCAAACTCCCTAGCAAACTCTAGTTGCATACAACTCGGGCCTTTTGTTGTTGGGATTTTGAAGAGCGATCTGTCCACGAGCATACATACGCCCTAGACATGGAC
ArgGlnLeuProAsnSerProSerLysLeuLeuValAlaTyrAsnArgGlyLeuCysValLeuTrpAspPheGluSerAlaSerValGlnArgAlaTyrIleAlaProArgHisGlyGln

1081 AGAGCGTTGGTCTTACAGTGAACCTCGAAGGTTGCAAGCTTACCTGACCGTATCCAGTGTGCTATACGCCCATTCAGGATACAGTAACCCAGACCGCGCTCGAATGTTAAATATTGTCG
SerValGlyLeuValLeuIleAsnPheGluGlySerGluArgTyrLeuThrIleHisGlyAlaIleArgTyrThrIleTrpSerIleAspAsnProGluProLysLeuValAsnThrValPro

1201 CTTATGGACCTGTCCATCGCAAAAGCATTAATCGACTGTACAAGGCGCAAGAGTCCAAAGCATGTAATGTTTTCGGGGCGGATCGCCACGGTCACGCAAAATGGGTGATCAAAATGGT
TyrGlyProAspProAspLysSerIleAsnArgLeuTyrLysGlyLysArgArgSerAsnAlaValPheSerGlyGlyMetProAlaAlaAsnGlyAspHisAsnCysVal

1321 TGTCGGTTACCGCCAGCGATGGACACAAAGTGTGCTTGACCTTTACGTCTAAAGTGATTGACATTTTGTGACCTTTGAAAATAGTAGAGATGTCGCTGAAGTTCTTGTGTACTACTTG
SerValLeuIleAsnArgAspGlyHisLysValCysLeuLysPheThrSerLysValIleAspIlePheValThrGlyAlaAsnSerArgAspValAlaGluValLeuValValLeuLeuGlu

1441 AAGAGGAACCTCGCGCTTACGATCTTACTGACCCCTAATTTTGTGCTACAAAAGGCCCATCTTCACTCTGTCCTGATGCATCAGCTGTGCTTGCATTTGCAATTTACTGCTTCTGACATCGTAC
GluGluLeuCysAlaTyrAspLeuThrAspProAsnIleCysAlaThrLysAlaIleArgProThrLysHisSerValHisAlaSerAlaValThrCysAsnThrLeuAlaSerGluLeuValGln

1561 AGTCGGTATATGAAGATTAATTAAGAGCTGGAGATGAACAGACATTCGACTATGACAATTAAGTCGGCTTACGATCGGGTACTCTCCGGATAACTAGAGAATCTCTGAGAAGAG
SerValTyrGluSerIleLeuArgAlaGlyAspGluLysAsnIleAspLysThrSerAsnIleSerProIleThrGlyGlyThrLeuProAsnLeuGluGluSerValGluGluAsp

1681 ACGCGCATTAAGCTTTTATGAGATTTTGTAACTGGTCACGAAGATGGTTCGTGTTAAATTTTGGGACTGCATGGAGTGTGCTTAAACCAATTTATAATTTTAAACATAGCAGCATTTTGG
AlaThrLysLeuTyrGluIleLeuLeuThrGlyHisGluAspGlySerValLysPheTrpAspCysThrGlyValLeuLeuLysProIleTyrAsnPheLysThrSerSerIlePheGly

1801 GAATGAGTCTAGACCTCCGAGATGACGACGTCGATCATCTAGTCGGGAACAAGTCGATGAAGGAAACCGCCATTTCCGAATACAGGACTTTTGTATCCTTATCAGATGACCTCGTT
SerGluSerAspPheArgAspAspAlaAlaAlaAspMetSerAlaGluGlnValLeuAspGluGlyGluProPheArgLysSerGlyLeuPheAspProTyrSerAspAspProArgLeu

1921 TAGCAATGAGAAGAAATGCAATTCGTGCCCAAAACCGGACCACTTATTGTTGGTGACAGCGGGCAAACTAGTTTATAGCGCACTCATAGACTACCCGAAAAAGGTCTTTAAAAATACA
AlaSerGluGluAsnAlaPheCysThrThrGlyGlnLeuValGlnValGlyThrAlaGlyGlnIleValGlyAlaAspPheIleAspLeuProGluLysValSerLysLysThrTyrlle

2041 TTTCAATGAATTTGGTCAGCGATCGTGATGGATTTGTGTGGAAGGTCACGATCAGTTAAACGTGCGATGCAACTTTATAGACGGAGAAGCAATTCCTACACGGAAACGTGGTGTAATA
SerMetAsnLeuValSerAspArgAspGlyPheValTrpLysGlyHisAspGlnLeuAsnValArgSerAsnLeuLeuGlyGluAlaIleProThrThrIleArgGlyValAsnIle

2161 TATCGGAGTACTGCAATGTTTTCGGCCAGCAGCATCACTGTCGACATCGGCACTGCAAGCAAGCTGGGGCGTAGTCTCTGGTGGAGCTGCGACGGCTGATCTTCTTGCATCTCAAAACT
SerGlyValLeuGlnValLeuProProAlaSerIleThrCysMetAlaLeuGluAlaSerTrpGlyLeuValSerGlyGlyThrAlaHisGlyLeuValLeuPheAspPheLysAsnPhe

2281 TTTGTCAGATTTACTCGCTGCATCTTAAACCCCAATGATCTTACGTGGACGAGGAGCAGCTGCTCGTCCGAAGTCTTTAAGAAGTCTTATGAGGAGTGATTTAGAAGCTTCGCA
ValProAlaPheHisArgCysThrLeuAsnProAsnAspPheThrThrGlyAlaGlyLeuArgArgLysPheThrLysLysSerLeuArgGluSerPheArgSerPheAlaArg

2401 GGGTCGATCAGCAACACCAACGACGCAATCAAGTACCAACAACGCTGGAAGCAAGACCGCTCGAGAGGCAATAGAGGCTCGTTGTGCGATGACGGGCTAGGATCCATGTCGCT
ValAspGlnThrLysThrAsnGlnSerAsnGlnValProThrThrLeuGluAlaArgProValGluArgGlnIleGluAlaArgCysAlaAspAspGlyLeuGlySerMetValArgCys

2521 GTTTACTTTTGGCAAACTATGTTACTAATGTCAACATACTGCTGCCATCTTTGTGGTTCAGCAACAATTCGCACTAGCTCTCGGTTTCTCTGCTATTTGCCACACCGCAGCAGCC
LeuLeuPheAlaLysThrTyrValThrValAsnGlnIleThrSerProSerLeuTrpSerAlaThrAsnAlaThrSerThrValSerValPheLeuLeuHisProArgAlaGlnThrAla

2641 CGGCACTCGCGCTCCGTCGGGCAAGTGGCAATGCACACCACTATGCCCGCGGAATTTTCGCGCAGCTGTGTAAGAAATAACAATTAACAACTCGTCTGCTGGTGGTATTTCTTA
AlaThrAlaValProCysProGluAlaSerGlyLysAlaSerGluProHisMetThrProArgArgIleSerAlaGlnLeuAlaValSerAlaGlnLeuLysHisArgAlaProValValGlyIleSerIle

2761 TTTTGTACAGGCGGTAGCCCTGTCGATCAGCTGAACCGCGGTGAAACGGGAGTCCACCGCATCGTGATCTATTGCTTCCGAGGAACAGTTCAGAGTGTTTTCACTTCGCCAATCAA
PheAspGlnAlaGlySerProValAlaAspGlnLeuAsnAlaGlyGluAsnGlySerProProHisArgValLeuIleAlaSerGluGluGlnPheLysValPheSerLeuProGlnLeuLys

2881 AGCGGATTAACAATTAAGCTTACCGTACAAGAGGTGCTCGGATTCGCGGCATCCATTTGCTTGGTTCGTTAGTGTGCGATATCCCGGAAACCTGCAGAGATGCACGGTGTAGCC
ProIleAsnLysTyrLysLeuThrAlaSerAlaGlnGluAlaIleArgArgIleHisPheGlySerPheSerHisGlySerProGluThrGlnSerMetHisGlyCysSerPro

3001 CAATTAAGTCCAGCGCTTCACATCGGCTGGAGAGCGGGATCTCAATATCAGTGGAGCTTGGCTGTAAAGTCGGGATGATATATAAGCAACAGCATGTATATGTTAAACGAATATGG
ThrLysSerThrArgSerHisGlyAspGlyGluAlaAspProAsnIleGlySerLeuAlaValSerArgGlyAspValTyrAsnGluThrAlaLeuIleCysLeuThrAsnMetGly

3121 GCGATATCATGGTTTATCAGTACCTGAAATTTAAAGACAGCTGAATGGCCGACAGCTGCGACGGGAAGACATTAATGGAGTTTCGTCACTTTGCTTTTCAAACTCTGGGAGAAGCATGT
AspIleMetValLeuSerValProGluLeuLysArgGlnLeuAsnAlaAlaAlaGluAlaAspGluAlaIleGlyValSerSerLysCysPheThrAsnSerGlyGluIleLeuTyr

3241 ATATGATGTCCTTTCTGAACTGACGGATCTTCTTACCGCAGCTCCAGAGCTGCAACCACTCGGCTGTTCCAGTAGAACATAGAATAAGAGCTTCTTGTGGAAGAAATG
MetMetSerSerSerGluLeuGlnArgIleAlaLeuAlaThrSerArgValValGlnProThrGlyValValProValGluProLeuGluAsnGluGluSerValLeuGluGluAsnAsp

3361 ATGCAAGAAATATAAGGAACCTTACGATCGTGTAGTGAAGTTGGAATACATAGAAATTAATAATCCATACGCAATTCATATGCAAGAGCTGCAAGAGAAACGTTGGAAGAAATA
AlaGluAsnAsnLysGluThrTyrTyrAlaCysAspGluValGlnLeuThrTyrGluLysLysArgProSerGlyIleSerGlyThrGluAlaGluGlnValGlyArgAsnSer

3481 GTGTTACGAAAGTTAATGGAGTCAACATTTCAAATTCACCTTAATCAAGCTAACGAGACTATCAGCAGCTCTATTGGCGATATTACCGTGACTCGGTGCGGACCATTTAAATATGACGA
ValGlnGlnValGlyValGlnIleGlyValAsnIleGlySerAsnSerProGlnAlaGlnGluThrIleSerSerIleGlyAspIleThrValAspSerThrArgAspHisLeuAsnMetThrThr

3601 CCACCATCTTGTGTTTAAATACAGAGAAACATTTGGTCGCTTCTCTGATCAGCAGCAACCAACGAAGGCACTACCGTAAACATGAGTGAAATTCCTTAAATTAATTAATTT
ThrThrLeuCysSerIleAsnThrGluGluThrIleGlyArgLeuSerValLeuSerThrGlnThrAsnLysGlySerThrThrValAsnMetSerGluIleProAsnIleAsnIleSer

3721 CTAATTTAGGAGCTTGAATCGAAAGAAATACGACGGAAACGAGTACTAGTTCGTTGTAATTAATCTATAATCAAACTTCTTACGAAAAACGAAACGAGGACCAAAATAG
AsnLeuGlu

Fig. 2. Nucleotide sequence of the pE7-9 clone and deduced amino acid sequence. Nucleotides are numbered from the beginning of the poly(dG) tail (added to the 5' end of the cDNA before its incorporation into the phage lambda gt10 vector). Amino acid residues are indicated below the nucleotide sequence, beginning from the first AUG codon in-phase with the open reading frame. A second AUG codon is found at position 538. A possible transmembrane signal peptide similar to a consensus sequence (Perlman and Halvorson, 1983) is boxed. Potential glycosylation sites (AsnXSer and AsnXThr) are indicated by black dots. A polyadenylation signal (Proudfoot and Brownlee, 1976) is found at position 5228.

Fig. 3. Pattern of charged amino acid residues and hydropathy profile of the deduced amino acid sequence from the open reading frame of the pcE7-9 clone. Arrows denote internal tandem repeats found in the deduced amino acid sequence. Repeats are numbered according to their relative similarity. Repeats 11 and 12 differ from all the other repeats. Potential N-glycosylation sites (open circles) and cysteine residues (black dots) are indicated.

Fig. 4. Internal amino acid sequence similarities within the deduced protein sequence. Potential N-glycosylation sites are marked with thick lines. Amino acid residues identical in at least two positions are boxed. Conservative amino acid substitutions are indicated by dots. A potential cell-contact site RGDV, as described in fibronectin (Hynes, 1985; Naidet *et al.*, 1987), is present in repeat 4. Conserved amino acid residues including conservative substitutions are indicated. Repeats are numbered as in Figure 3.

eluted from the column. The antibodies were tested on Western blots and shown to bind exclusively to the *Drosophila* part of either fusion protein (Figure 7c).

Western blots containing electrophoretically separated protein extracts from wild-type and homozygous *l(2) gl* mutant larvae were tested with the affinity-purified serum. A protein of ~130 kd in size is labeled in wild-type protein and absent in the null mutant *l(2) glDV275* (Figure 7d). This mutation has been shown recently to comprise an internal deletion of ~8 kb from the 5' half of the gene (Lützelschwab *et al.*, 1986). A protein

Fig. 5. Short consensus sequence of the internally repeated sequences from the deduced 130-kd protein of the *l(2) gl* gene. The number indicates the position of the first amino acid residue in the deduced amino acid sequence in Figure 2. Sequences are aligned according to their respective cysteine positions. Relative to cysteine (position 0) are negatively charged amino acid residues (at position -8 to -11 and +5 to +7) and asparagine residues (at -6 to -8 and +3 to +5). Also included are similar regions from the deduced protein sequences of chicken N-CAM (Hemperly *et al.*, 1986); chicken integrin, a membrane glycoprotein (Tamkun *et al.*, 1986); and *Dictyostelium* contact site A protein (Noegel *et al.*, 1986).

of 130 kd in size is also observed in protein extracts of wild-type embryos (Klämbt and Schmidt, 1986).

Recently, fusion proteins from the plasmids pPTB5, pPTH2 and pPTBa1 (Figure 1) were used to raise antibodies in rabbits.

Fig. 6. Internal amino acid sequence similarities within the deduced protein sequence of the contact site A protein of *Dictyostelium* (Noegel *et al.*, 1986). The deduced protein sequence of the csA protein was aligned according to the observed spacing of potential glycosylation sites, serine and aliphatic amino acid residues which is similar in some of the *Drosophila* repeats (Figure 4).

Fig. 7. Characterization of antiserum against a bacterial fusion protein and affinity purification of 130-kd protein-specific antibodies. (a) Extracts from the bacterial strains containing the plasmids pPGH5 (g+), pUR291 (g-), pPTH6 (t+) and pATH10 (t-) electrophoretically separated on a SDS-polyacrylamide gel and stained with Coomassie blue. Black arrowheads indicate the position of the fusion proteins, white arrowheads indicate the position of proteins expressed by the vector plasmids. (b) Western blots of bacterial proteins labeled with rabbit antiserum against the 67-kd fusion protein. Extracts from bacterial strains containing the plasmids pPGH5 (g+), pUR291 (g-), pPTH5 (t+) and pATH10 (t-) were electrophoretically separated as in (a), transferred onto nitrocellulose filters and then treated with total antiserum. Bound antibodies were visualized histochemically with a second antibody coupled with horseradish peroxidase. (c) Affinity-purified antiserum against the *l (2) gl/trpE* fusion protein. Western blot of lanes t+ and t- (as in b), labeled with antiserum eluted from a column containing covalently bound *l (2) gl/lacZ* fusion protein. (d) Western blot of *Drosophila* protein extracts from brains of third instar larvae, treated with affinity-purified antiserum. Left are proteins from wild-type brains, right are proteins of homozygous mutant brains from the strain *l (2) glDV275* (Lützelschwab *et al.*, 1986). Similar results were obtained with protein extracts from whole larvae.

The antisera obtained from these fusion proteins recognize a 130-kd protein in *Drosophila* wild-type protein extracts.

Subcellular localization

To determine the location of the 130-kd protein in *Drosophila* cells we analysed antibody-treated sections of embryos with indirect immunofluorescence techniques.

In embryonic sections cellular membranes of presumptive larval cells were labeled (Figure 8). In addition, the axon bundles of the neuropile are labeled, whereas cell bodies of neural tissues showed much less antigen (Figure 8). Since axon bundles in the neuropile consist mainly of membranes, the strong label in these

structures is probably due to the localization of the protein at the cellular membrane or in the intercellular matrix.

Discussion

We report two independent lines of evidence which provide information on one of the protein products from the *l (2) gl* gene. First, the nucleotide sequence of the ~5.3-kb cDNA fragment revealed an open reading frame predicting a 130-kd protein. Second, we have raised antibodies against fusion proteins which recognize a 130-kd protein in embryonic and larval *Drosophila* protein extracts. The protein is absent in extracts from homo-

Fig. 8. Antibody binding to parasagittal section of a wild-type embryo. Sections were treated with affinity-purified antiserum (see Figure 7) and washed with phosphate-buffered saline. A second, FITC-coupled antibody was applied and after several washes the sections were overlaid with 4% *n*-propyl gallate in glycerol and inspected for fluorescence under appropriate u.v. light. The section is slightly overstained to show the labeling of the neuropile. Abbreviations: mg, midgut; np, neuropile; ph, pharynx; pv, proventriculus; sup, supraesophageal ganglion; vc, ventral cord.

zygous mutant larvae and in tissue sections of mutant embryos.

The presence of a transmembrane signal peptide in the predicted protein sequence and the detection of the 130-kd protein at the cellular membrane in tissue sections suggest that the protein exerts its function in the extracellular space. The protein could be anchored into the cellular membrane by the transmembrane signal peptide (Spiess and Lodish, 1986) or be bound to other components of the cellular membrane or extracellular matrix.

The phenotypic effects of *l (2) gl* mutations are due to the absence of a functional *l (2) gl* protein in homozygous mutant tissues. This visibly affects the structural integrity of cells in the optic centres of the brain and imaginal discs (Gateff, 1978a,b). When implanted into the abdomen of adult female flies, mutant tissues grow in an abnormal, disperse fashion, suggesting that cell adhesion properties have changed in these cells (Gateff, 1978a,b).

In the deduced sequence of the 130-kd protein from the *l (2) gl* gene we could identify internal repeats of similar amino acid sequences. Within the repeats conserved features appear to be similar to protein sequences of the extracellular part of the chicken N-CAM protein (Hemperly *et al.*, 1986) and the contact site A protein of *Dictyostelium* (Noegel *et al.*, 1986). Using potential glycosylation sites as reference points, four internal repeats of similar sequences are detected in the extracellular domain of the csA protein. Some of the structural features, such as serine and threonine residues, charged amino acids and potential glycosylation sites, which are regularly spaced within the *Drosophila* repeats, are at a similar position in the csA repeats (Figure 6). Although the functional significance of the conserved features in those proteins remains to be determined, the observed structural similarity between the 130-kd protein of the *l (2) gl* gene and two proteins which are functionally described as cell adhesion proteins (Edelmann, 1983; Gerisch, 1986; Rutishauser, 1985) suggests similar functional properties for these proteins. In summary, the phenotypic effects caused by the *l (2) gl* mutation, the location of the 130-kd protein at the cellular membrane and amino

acid sequence similarities to proteins with reported cell adhesion properties are all consistent with the notion that this protein is involved in cell-cell interaction.

Materials and methods

cDNA fragments

cDNA clones were isolated from a lambda phage gt10 library (kindly provided by L.Kauvar and T.Kornberg) containing cDNA fragments synthesized from poly(A)⁺ RNA of different developmental stages (Poole *et al.*, 1985). Independent isolates were obtained from libraries constructed from poly(A)⁺ RNA of the blastoderm and third instar larvae stages. For the screen genomic DNA fragments (Mechler *et al.*, 1985; Lützeltschwab *et al.*, 1986), originally provided by B.Mechler, were used as probes. About 40 recombinant phages containing cDNA fragments from the *l (2) gl* locus were isolated. Three phages (cE7-9, cE7-19 and cE7-20) contained fragments > 5 kb. For further analysis the cDNA fragments were subcloned into the plasmid pbs-neo (Steller and Pirrotta, 1985).

DNA sequencing

DNA sequencing was performed by the chemical degradation method (Maxam and Gilbert, 1980). The *EcoRI* cDNA fragment from the phage cE7-9 was subcloned into the pbs-neo vector (Steller and Pirrotta, 1985) and deletion clones were generated with the use of deoxyribonuclease in the presence of Mn²⁺ (Frischauf *et al.*, 1980). At least 20 deletion clones with breakpoints approximately every 250 bp were analysed and ~500 bp were sequenced from one end. Thus each part of the clone was sequenced at least twice. In addition, genomic DNA fragments covering most of the cDNA sequences were sequenced in both directions.

Protein scan

Compiled protein sequences of the Dayhoff/Doolittle libraries (Kyte and Doolittle, 1982; Orcutt *et al.*, 1982) were scanned with a short consensus sequence found within the repeated sequences (and elsewhere) in the deduced amino acid sequence of the 130-kd protein, as follows. A characteristic spacing of negatively charged amino acid residues (Asp and Glu) and Asn with regard to cysteine residue was observed at six different locations in the deduced protein sequence. An algorithm to scan protein sequences of the libraries allowed Asp and Glu residues at positions -11 to -8 and +5 to +7 and Asn at position -8 to -6 and +3 to +5 relative to a cysteine residue. All other positions were undefined except that cysteine residues were allowed only at position 0.

Seven of the 12 sequences identified in this scan, integrin, variable surface glycoprotein 3, contact site A protein, neural cell adhesion molecule, glycosylated surface protein, alpha chain from factor XII and glutathione peroxidase, are pro-

teins located at or outside the cellular membrane. Four others are viral proteins and two more (spectrin and tubulin) are intracellular proteins.

Affinity-purified antibodies

Two types of expression vectors were used to synthesize fusion proteins in bacteria. The pATH expression vectors (kindly provided by T.J.Körner, New York, unpublished data) contain a truncated coding sequence of the bacterial gene *trpE* from the tryptophan operon. A polylinker located adjacent to the 3' end of the coding sequence allows the insertion of foreign DNA fragments in the desired orientation in all three possible reading frames. The pUR expression vectors contain an almost complete coding sequence of the *lacZ* gene from the β -galactosidase operon with a polylinker at the 3' end of the coding sequences (Rüther and Müller-Hill, 1983). We have ligated a 2.2-kb *Hind*III fragment from the pE7-9 cDNA clone (see Figure 1b) into the *Hind*III site of the expression vectors pATH10 and pUR291. From the cDNA sequence both constructs code for an additional 0.9 kb of open reading frame sequences. This would amount to ~30 kd of additional protein expressed in both constructs. A 67-kd fusion protein was synthesized from pPTH6 compared with 37-kd for the *trpE* vector protein alone. Similarly, a fusion protein of 145 kd from the pGTH5 plasmid was detected, compared with 116 kd for the *lacZ* β -galactosidase protein.

To obtain preparative amounts of the fusion protein, the bacterial strain C600 containing the plasmid pPTH6 was grown to midlog phase and the tryptophan promoter was induced with 3, β -indoleacrylic acid. About 100 μ g of gel-purified fusion protein was injected into rabbits, followed after 6 weeks with 100 ng as booster. After another 10 days the serum was tested on Western blots. The serum showed no cross-reaction to β -galactosidase. The total antiserum (5 ml) against the *l* (2) *gl*/*trpE* fusion protein was purified in a single step using a (10 ml) Sepharose 4B column with covalently bound *l* (2) *gl*/*lacZ* fusion protein by incubating at 6°C overnight. After washing, the specifically bound antibodies were eluted with 3 M MgCl₂.

Acknowledgements

We thank L.Kauvar and T.Kornberg for use of their cDNA library and Dr Körner for providing the pATH expression vectors prior to publication. We also thank B.Mechler for communicating unpublished results of the DNA transformation experiments and for a helpful critique of errors in our own initial sequence, based on his genomic sequences. Arno Uhlmann was instrumental in our computer work. We also thank R.Cassada, H.Kössel and K.Sander for critically reading the manuscript. C.K. is a recipient of a GFG fellowship. This work was supported by the Deutsche Forschungsgemeinschaft SFB 31.

References

- Edelmann, G.M. (1983) *Science*, **219**, 450–457.
 Frischauf, A., Garoff, H. and Lehrach, H. (1980) *Nucleic Acids Res.*, **8**, 5541–5548.
 Garcia-Bellido, A. (1966) *Dev. Biol.*, **14**, 278–306.
 Garcia-Bellido, A. (1975) In *Cell Patterning. Ciba Foundation Symposium*. Elsevier, Amsterdam, Vol. 29, pp. 161–182.
 Gateff, E. (1978a) *Science*, **200**, 1448–1452.
 Gateff, E. (1978b) In Ashburner, M. and Wright, T.R.F. (eds), *Genetics and Biology of Drosophila*. Academic Press, NY, Vol. 2b, pp. 181–275.
 Gateff, E. and Schneiderman, H.A. (1967) *Am. Zool.*, **7**, 760.
 Gateff, E. and Schneiderman, H.A. (1969) *Natl. Cancer Inst. Monogr.*, **31**, 365–397.
 Gateff, E. and Schneiderman, H.A. (1974) *Wilhelm Roux's Arch.*, **176**, 23–65.
 Gauger, A., Fehon, R.G. and Schubiger, G. (1985) *Nature*, **313**, 395–397.
 Gehring, W.J. (1985) *Cell*, **40**, 3–5.
 Gerisch, G.J. (1986) *J. Cell Sci.*, Suppl. **4**, 201–219.
 Hemperly, J.J., Murray, B.A., Edelman, G.M. and Cunningham, B.A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3037–3041.
 Hynes, R.O. (1985) *Annu. Rev. Cell Biol.*, **1**, 67–90.
 Inouye, M. and Halegoua, S. (1980) *CRC Crit. Rev. Biochem.*, **7**, 339–371.
 Klämbt, C. and Schmidt, O. (1986) *EMBO J.*, **5**, 2955–2961.
 Kozak, M. (1984) *Nucleic Acids Res.*, **12**, 857–864.
 Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.*, **157**, 105–113.
 Lewis, E.B. (1978) *Nature*, **276**, 565–570.
 Lützelshwab, R., Müller, B., Wälder, B., Schmidt, O., Fübress, R. and Mechler, B. (1986) *Mol. Gen. Genet.*, **204**, 58–64.
 Mahowald, A.P. and Hardy, P.A. (1985) *Annu. Rev. Genet.*, **19**, 149–177.
 Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–561.
 Mechler, B., McGinnis, W. and Gehring, W.J. (1985) *EMBO J.*, **4**, 1551–1557.
 Naidet, C., Semeriva, M., Yamada, K.M. and Thiery, J.P. (1987) *Nature*, **325**, 348–350.
 Noegel, A., Gerisch, G., Stadler, J. and Westphal, M. (1986) *EMBO J.*, **5**, 1473–1476.
 Nüsslein-Volhard, C. and Wieschaus, E. (1980) *Nature*, **287**, 795–801.

- Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) *Database Search Program: SEARCH*. NBR Report 820503-08710. National Biomedical Research Foundation, Washington, DC.
 Perlman, D. and Halvorson, H.O. (1983) *J. Mol. Biol.*, **167**, 391–396.
 Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) *Cell*, **40**, 37–43.
 Postlethwait, J.H. and Schneiderman, H.A. (1971) *Dev. Biol.*, **24**, 477–519.
 Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature*, **263**, 211–213.
 Rüther, U. and Müller-Hill, B. (1983) *EMBO J.*, **2**, 1791–1794.
 Rutishauser, U. (1984) *Nature*, **310**, 549–554.
 Spiess, M. and Lodish, H.F. (1986) *Cell*, **44**, 177–183.
 Steiner, D.F., Quinn, P.S., Patzelt, C., Chan, S.J., Marsh, J. and Tager, H.S. (1980) *J. Cell Biol.*, **4**, 175–202.
 Steller, H. and Pirrotta, V. (1984) *EMBO J.*, **4**, 167–171.
 Tamkun, J.W., DeSimone, D.W., Fonda, D., Patel, R.S., Buck, C., Horwitz, A.T. and Hynes, R.O. (1986) *Cell*, **46**, 271–282.
 White, R.A.H. and Wilcox, M. (1984) *Cell*, **39**, 163–171.

Received on January 15, 1987; revised on March 17, 1987