

A novel cDNA from *Drosophila* encoding a protein with similarity to mammalian cysteine-rich secretory proteins, wasp venom antigen 5, and plant group 1 pathogenesis-related proteins

Michael C. Schreiber, J. Colleen Karlo, Gae E. Kovalick *

Department of Zoology, 212 Pearson Hall, Miami University, Oxford, OH 45056, USA

Received 8 August 1996; received in revised form 26 November 1996; accepted 29 November 1996; Received by D. Schlessinger

Abstract

The CAP protein family is made up of a group of secreted proteins that share sequence similarity. Members of this family are found in animals, plants, and fungi, and their shared sequence similarity suggests that members share a common, but as yet unknown, molecular function. As a first step in defining the function of CAP family proteins, an 878 bp partial cDNA encoding a novel member of the CAP family was cloned by the polymerase chain reaction (PCR) from total RNA of adult *Drosophila*. The cDNA contained the complete coding sequence for a protein 256 amino acids in length, as well as the complete 3' untranslated region (UTR) and a portion of the 5' UTR. The protein, named Antigen 5-related (Agr), was most similar in sequence to antigen 5 (Ag5), a CAP family member found in social wasps and ants. The corresponding *Agr* RNA is about 1 kb in length and is present at all stages of development, with highest levels observed in adults. *Agr* RNA is transcribed from a single gene that is located within region 12F of the X chromosome. The identification of *Agr* in *Drosophila* expands the number of known CAP family members to well over four dozen. Further studies of *Agr* and the gene which encodes this protein using the *Drosophila* model system may help provide important insight into the molecular functioning of this little known, but increasingly significant protein family. © 1997 Elsevier Science B.V.

Keywords: *Drosophila* DNA sequence; Extracellular proteins; Secreted proteins; Polymerase chain reaction

1. Introduction

The mammalian cysteine-rich secretory proteins (CRISPs), the antigen 5 (Ag5) proteins of social wasps and ants, and the group 1 pathogenesis-related (PR-1) proteins of plants have been known for some time to compose a family of secreted proteins that share amino acid sequence similarity (Fang et al., 1988; Lu et al., 1993). CRISPs have been identified in both rodents and

humans; their expression is androgen-regulated and generally localized to the male reproductive tract (Brooks et al., 1986; Haendler et al., 1993; Hayashi et al., 1996). Some of these proteins have been shown to bind sperm and to be required for sperm-egg fusion during fertilization (Cuasnicú et al., 1984; Hayashi et al., 1996). Ag5 is found in the venoms of social wasps and fire ants (King et al., 1978) and is probably involved in producing an intense, localized inflammatory reaction in mammals that have been stung. The PR-1 proteins are synthesized by plant cells in response to various stresses, including pathogen infection, and have recently been shown to have anti-fungal activity in both in vitro and in vivo assays (Niderman et al., 1995).

Although the physiological processes in which the original members of the CRISP/Ag5/PR-1 (CAP) family participate have been outlined, their molecular action is unknown, and the functional relationship between the different member groups remains obscure. In recent years, a variety of other proteins with sequence similarity to known CAP family members have been identified,

* Corresponding author. Tel. +1 513 5293128; Fax +1 513 5296900; e-mail: kovalige@muohio.edu

Abbreviations: Ag5, antigen 5; Agr, antigen 5-related; AP, alkaline phosphatase; bp, base pair(s); CAP family, proteins related by amino acid sequence to CRISPs, Ag5, and PR-1 proteins; cDNA, DNA complementary to RNA; CRISP, cysteine-rich secretory protein; DIG, digoxigenin; GSP, gene-specific primer; kb, kilobase(s) or 1000 bp; mRNA, messenger RNA; nt, nucleotide(s); PCR, polymerase chain reaction; PR, pathogenesis-related; PR-1 protein, group 1 pathogenesis-related protein; R, purine; RACE, rapid amplification of cDNA ends; SGP28, specific granule protein 28; UTR, untranslated region; X, any deoxynucleotide; Y, pyrimidine.

suggesting that the CAP family proteins are more widely distributed than previously suspected. Characterization of these new CAP family proteins have the potential to provide increasing insight into the molecular role of CAP proteins. Novel members include proteins from the fungi *Saccharomyces cerevisiae* (Miosga et al., 1995) and *Schizophyllum commune* (Schuren et al., 1993), glioma PR protein from human tumor cells and macrophages (Murphy et al., 1995), specific granule protein 28 (SGP28) from human neutrophils (Kjeldsen et al., 1996), and helothermine, a protein found in the venom of the Mexican beaded lizard (Morrisette et al., 1995). The function of these newly identified proteins is for the most part unknown; however, based on its intracellular location within the specific granules, SGP28 has been hypothesized to function either as an antimicrobial protein or as a novel matrix protease (Kjeldsen et al., 1996). Direct analysis of protein function has so far yielded little information on the molecular role of CAP family proteins. We have decided to use a genetic and molecular approach to the problem of CAP family function. The *Drosophila* model system is a particularly powerful system for the genetic and molecular analysis of protein function. We describe here the first step in our approach: the isolation and sequence of a cDNA encoding a member of the CAP family, named Antigen 5-related (*Agr*), from *Drosophila melanogaster*. We also examine the developmental expression of the RNA and the localization of the *Agr* gene. The data described here will form the foundation for further studies into the structure, expression, and function of the *Agr* gene and the protein it encodes. This, in turn, will help increase our understanding of the biological function of other CAP family proteins.

2. Experimental and discussion

2.1. Isolation, cloning, and sequencing of *Agr* cDNA

While attempting to clone an unrelated cDNA by PCR, we fortuitously amplified a cDNA fragment from *Drosophila melanogaster* that encoded a protein with sequence similarity to wasp venom Ag5. The method used for cloning was a variation of the 3' RACE (rapid amplification of cDNA ends) technique of Frohman et al. (1988). The template for the PCR reaction was cDNA synthesized by reverse transcriptase from adult *Drosophila* total RNA. The primer (GGCCACGCGTCGACTAGTAC(T)₁₇) used for cDNA synthesis was complementary to the poly(A) tail of messenger RNA (mRNA), but also had a noncomplementary 5' sequence that contained recognition sites for the restriction enzymes *Mlu*I, *Sal*I, and *Spe*I and a half-site for *Not*I (Gibco BRL). Two oligonucleotide primers were used for the amplification reaction: a universal primer that

corresponded to the noncomplementary sequence at the 5' end of the primer used for cDNA synthesis (see above; Gibco BRL), and a degenerate 'gene-specific' primer (AARAAYTAYCARRTXTGG, where R = A/G, Y = T/C, and X = any deoxynucleotide) that was complementary to a conserved region within the cDNA we were attempting to amplify. Thirty-five cycles of 1 min at 94°C, 1.5 min at 45°C, and 2 min at 72°C were carried out. Amplified fragments were then separated by size using agarose gel electrophoresis. A product of approximately 880 bp was purified from the gel matrix and ligated into T-tailed (Holton and Graham, 1991) pBluescript II vector (Stratagene) for sequencing.

Two independent clones obtained by PCR amplification were completely sequenced using the dideoxy method (Sanger et al., 1977) modified for nonradioactive detection of the DNA fragments. Minor sequence discrepancies between the two clones were resolved by partially sequencing two other PCR-amplified clones. Because of the degeneracy of the gene-specific primer (GSP), the annealing conditions during PCR were such that mispairings between the GSP and the template cDNA were possible. To verify the sequence of the cDNA near the GSP, genomic DNA near the annealing site of the GSP was also amplified by inverse PCR (Triglia et al., 1988) using non-degenerate primers and more stringent annealing conditions. The amplified product was then cloned and sequenced as described above. Although the GSP annealing site within all of the cDNAs corresponded to the sequence of the GSP, five mismatches were found between the GSP and the genomic sequence (Fig. 1). The consensus nucleotide sequence of the cDNA is shown in Fig. 1. The cDNA was amplified from both PCR primers and contained a polyadenylation signal and a poly(A) tail. The cDNA contained a long open reading frame, which began at the twenty-seventh nucleotide in the sequence and terminated at nucleotide 795, encoding a protein 256 amino acids in length. The genomic sequence contained an in-frame termination codon upstream of the first methionine codon. This termination codon was not found in the cDNA sequence. The presence of the termination codon suggests that the cDNA contained a complete protein coding region as well as a portion of the 5' UTR. However, the possibility that the in-frame stop codon is part of an upstream intron and thus spliced out of the processed RNA cannot as yet be discounted. The cDNA also has a very short 3' UTR of 36 bp.

2.2. Sequence analysis

A nucleotide homology search through the GenBank and EMBL databases using the program FASTA (Pearson and Lipman, 1988) indicated that the cDNA represented a novel sequence. An analysis of codon frequency (data not shown) within the open reading

AARAAYTAYCARRXTGG		
GCTAGCAATCAGGTCTGGGAATAGAAATGCGCAACTTCGTGATCATATTAGCCTATCTC	<u>Bcl I</u>	60
	M R N F V I I F S L S L	12
TGGCACTTGGCATCGCTCCGCCACCGATTATTGCAAAAAGAGCTGCGGAAGCACCAAGA		120
A L G I A S A T D Y C K K S C G S T K N		32
ATCTGGGATGCGACAATAATGAGCCTGGGCTCAAGCTGTCCAGTGACGCCACCTTT		180
L G C D N N G A W A S S C P S D A T L L		52
TGACCTCTCCAGCGCTGAGAAGGATGCACTGGTGCCAGGACGAACGATATCGCAACC		240
T L S S A E K D A L V A R T N E Y R N H		72
ACATCGCCGGGACTGAATGCCAATCTGAGTGCCGCTGTGCAATGGCCACGATCAAGT		300
I A G G L N A N L S A A C R M A T I K W		92
GGAACGATGAATGGCTTACTTGGCAGTTTGAACGTGAAAGTTGTCAAATGAACACG		360
N D E L A Y L A S L N V K S C Q M K H D		112
ACGGCTGCCAATACGGATGCTTCTGACTGGTCTCGGAGAATCTGGCTGGATGGCT		420
G C H N T D A F D W S P E N L A W M G Y		132
ACTACAATCCGTAATGTTACACACTATCTAGAATGGGCGTCCGATGTGGTACGATG	<u>Xba I</u>	480
Y N P L N V T H Y L E W G V D M W Y D E		152
AGCGGTGTACACCAACAGGCGCTCATGATCGCTCGAATCAATGGTCCGG	<u>Cla I</u> <u>Ava II</u>	540
A V Y T K Q A Y I D A Y P S N Y N G P A		172
CCATTGGACATTTACGGTGTCTGTTCCGATCGGAATACGGAAGTGGTGTGCGCGG	<u>Pvu I</u>	600
I G H F T V L V A D R N T E V G C A A A		192
CCAGTACTCCGTGTCGGCCCAATCGTACAAGGCTTCTGCTGGCTGTAATATCGG		660
T Y S V S G Q S Y K A F L L A C N Y A A		212
CCACCAATGTTCTGGGCATCAAGATGTACAGTTCTGCTCAAGGCGGCGACAGTGTA		720
T N V L G I K M Y S S C S K A A S K C T		232
CCACCGTACCAATCCAAGTACAAGTATCTGTGACGCGCAAGGAGGAGTACAACGTAA	<u>Kpn I</u>	780
T G T N P K Y K Y L C S A K E E Y N V N		252
ACAACCTCTTCTATTGAGGCAATAATAGATGCTACGATGGTTCAATTATTTAAAAA	<u>Pvu I</u> <u>Dra I</u>	840
N L F Y *		256
AAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCC		879

Fig. 1. Consensus sequence of a partial *Agr* cDNA from *Drosophila*. The consensus nucleotide sequence is shown, along with the putative protein sequence of the longest open reading frame. The complete sequences of two *Agr* cDNAs amplified by PCR were determined and compared. Discrepancies in the nucleotide sequence found between the two clones were resolved by partial sequencing of two independent clones. With the exception of the first 15 nt, which were derived from genomic DNA, the sequence shown consists of the consensus sequence obtained from the four cDNAs. The sequence of the GSP, underscored with a dotted line, is shown above the genomic sequence. The sequence within the cDNA that is complementary to the universal primer is also underscored with a dotted line. The in-frame termination codon upstream from the initiator codon is underlined. The putative signal sequence is underscored with a double line. Recognition sites for restriction enzymes are indicated with the placement of the enzyme name above the site. The polyadenylation signal is shown in bold face type. The sequence data shown have been submitted to the EMBL/GenBank/DBJ Data Libraries under the accession number L49036.

frame indicated that codon usage was consistent with that of most moderately to highly expressed *Drosophila* coding sequences (Sharp and Lloyd, 1993).

Using the program TFasta, a homology search of the GenBank, EMBL, and PIR sequence databases was made with the amino acid sequence encoded by the cDNA. Significant homology was found to the Ag5 protein of various wasp species. The amount of identity

between the *Drosophila* cDNA and Ag5 varied from 26 to 28%, depending on the species, over a region 202 amino acids long (Fig. 2). When conservative amino acid substitutions were also included, the amount of similarity was estimated to be 47–51%. Because of the similarity between the *Drosophila* sequence and Ag5, we termed the *Drosophila* protein Antigen 5-related (*Agr*).

The Ag5 proteins are members of a larger family of proteins related by sequence (the CAP family). A comparison of the *Drosophila* *Agr* amino acid sequence with that of other CAP family proteins is shown in Fig. 2. The degree of identity between the various CAP family proteins shown varies from 24 to 57% within the conserved region (extending in *Agr* from amino acid 66–221). Within this region, the degree of identity between *Agr* and the other CAP family proteins listed is 25% (pSC7, SGP28, helothermine) and 27% (PRP1b). There are 18 residues that are invariant between the six proteins; three of these conserved amino acids are cysteines. These cysteines have been shown to form intramolecular disulfide bonds in Ag5 (King et al., 1990). All CAP family proteins are secreted and are initially synthesized with a signal sequence at the amino terminus. The amino terminal end of the *Agr* protein contains a 19 amino acid sequence that could function as a signal sequence (Fig. 1). Thus, *Agr* itself may be secreted. Nishikawa et al. (1983) have developed a method of protein sequence analysis that allows the classification of proteins into four groups with some degree of accuracy: intracellular nonenzymes, intracellular enzymes, extracellular nonenzymes, and extracellular enzymes. Analysis of the amino acid composition of the *Agr* protein suggests that it can be categorized as an extracellular enzyme (data not shown). The classification of *Agr* as an extracellular protein is consistent with the other evidence that *Agr* is secreted and functions outside the cell.

2.3. Northern analysis

The size of the *Agr* transcript and its developmental expression were determined by Northern blotting of total RNA extracted at various stages of *Drosophila* development and by hybridization of antisense *Agr* RNA to the blot (Fig. 3A). The presence of transcript at each stage of development was also assessed by reverse transcription-PCR (Fig. 3C). In Northern blots, the probe hybridized to a single transcript that was estimated from several experiments to be approximately 1000 nt. The transcript was present at all developmental stages and in both males and females. Transcript levels were highest in adults; RNA levels were much lower in all of the preadult stages. Comparison of the size of the full length transcript with that of the cloned cDNA

Agr	--MRNFVIIF	SLSLALGIAS	A-----	TDYCKKSCGS	TKNLGCDDNG	AWASSCPSDA	--TLLTLSSA	EKDALVARTN	67
VVAg5	--MEISGLVY	LIIVVTIIDL	PYGKA----	NNYCKIKKCL	GGVHTACKYG	SLKPNCGNKV	VVSYGLTKQE	-KQDILKEHN	72
pSC7	--MKLTVILL	TAVLAASASP	AP-----	-VDVDA---R	APVALDSRSP	IDIDSRSDA	LANRAAPPQS	EIQDWLKAHN	66
PRP1b	MGFFLFSQMP	SFFLVSTLL	FLIISHSSHA	-----	-----	-----	-----QN	SQQDYLDHN	42
SGP28	--MTLFPVL-	-LFLVAGLLP	SFPANED--	-----	-----	-----	-----KDPA	FTALLTTQTQ	47
Helothermine	--MILLSLY-	-LCLAAML--	-----	-----	-----	-----	-----EASPK	LPGLMTSNPD	39
CONSENSUS	M L	L						HN	
Agr	ERYNHIAGGL	N-----ANLS	AACRMATIKW	NDELAYLASL	NVKS----	CQ MKHDGCHNTD	AFDW-SPENL	AWMGYYNPLN	137
VVAg5	DFRQKIARGL	ETRGNPGPOP	PAKNMKNLVW	NDELAYVAQV	WANQ----	CQ YGHD--TCRD	VAKYQVGQNV	-ALT-GSTAA	144
pSC7	NERAQ-----	-----	--HGAVALVW	NQTLSDKAAD	WASQ-----	CI WEHSN----	-----SGQNL	AAWF-SPQAN	114
PRP1b	TARADVG-----	-----	---VEPLTW	DNGVAAYAQN	YVSQLAADCN	LVHS--HG--	-----QYGENL	-AQG-S---G	92
SGP28	ELRRRAVS-----	-----P	PARNMLKMEW	NKEAANAQK	WANQ-----	CN YRHSNPKDR-	MTSLKCGENL	-YMS-SAPSS	108
Helothermine	NLRRIVE-----	-----P	TASNMLKMTW	SNKIAQNAQR	SANQ-----	CT LEHTSKEERT	IDGVECGENL	-FFS-SAPYT	101
CONSENSUS	R		A...M.....W.N.A.AQ	A.Q.....C.....H			GENL.....S		
Agr	VTHYLEWGV	MWYDEAVYTK	QAYIDAYPSN	YNGPAIGHFT	VLVADRNTVE	GCAAATYSVS	G-----QS	YKAFLLLACNY	210
VVAg5	KYDDPVKLVK	MWEDE--VK	DYNPKKKFSG	NDFLKTGHYT	QMVWANTKEV	GCGSIKIQE	KWHK-----	---HYLVLCNY	212
pSC7	KPMNISQGVG	GWNAE---EP	DYNTTTSYGA	-----GHWT	QVWKSSTVS	GCAAYSCPPG	TLGRKPTDPW	KTLLWYVCNY	185
PRP1b	DFMTAAKAVE	MWDE---KQ	YYDHDSNTCA	QGQV-CGHYT	QVWNRNSVRV	GCAVVKCNG	GY-----	---VVSVCNY	156
SGP28	W----SQAIQ	SWFDE---YN	DFDFGVGPKT	PNAV-VGHYT	QVWYSSYLV	GCGNAYCPNQ	KVLK-----	---YYYVCQY	171
Helothermine	W----SYAIQ	NWFDE---RK	YFRFNYGPTA	QNMV-IGHYT	QVWYRSYEL	GCAIAYCPDQ	PTYK-----	---YYQVCQY	164
CONSENSUS		V.....W.DE		GHYT.QVW.....V.GCA.....C				Y.VCNY	
Agr	-AATNVLGIK	MYSSCSKAAS	KCTTGTPNPKY	KYLCSAKEEY	NNVNLFFY*				256
VVAg5	GPSGNFMNEE	LYQTK*							227
pSC7	YRPGNVSPRD	KYYPINVQP*							204
PRP1b	DPPGNVIGQS	PY*							168
SGP28	CPAGNANR-	LYVPYEQGAP	CASCPDNCDD	GLCTNGCKYE	DLYSNCKSLK	LTLTCKHQLV	RDSCKASCNC	SNSIY*	245
Helothermine	CPGGNIRSR-	KYTPYSIGPP	CGDCPADCN	GLCTNPCKQN	DVYNNCPDLK	KQVGCCHPIM	KD-CMATCKC	LTEI*	236
CONSENSUS	..P.GN.....Y								

Fig. 2. Comparison of the putative Agr protein sequence with related sequences from other species. The protein sequence of Agr was derived from the longest open reading frame of the partial Agr cDNA. The antigen 5 sequence from the wasp *Vespa vulgaris* (VVAg5) was determined both by amino acid sequencing of the entire purified processed protein (Hoffman, 1993) and by nucleotide sequencing of the cDNA (Lu et al., 1993). The human specific granule protein 28 (SGP28) sequence was determined by partial amino acid sequencing of the protein and cDNA sequencing (Kjeldsen et al., 1996). Sequences of pSC7 from the fungus *Schizophyllum commune*, group 1 pathogenesis-related protein from tobacco (PRP1b), and helothermine from the Mexican beaded lizard were obtained from cDNAs (Schuren et al., 1993; Ohshima et al., 1990; Morrissette et al., 1995). Sequences were aligned using the Genetics Computer Group program PILEUP (Devereux et al., 1984), followed by manual editing to optimize alignments. Dashes represent gaps inserted to improve alignment. A consensus sequence is listed below the aligned sequences at certain positions and represents those amino acids which are conserved in four or more of the sequences. Areas where conserved amino acids are clustered are underscored within the consensus sequence by dotted lines.

fragments indicated that no more than 170 bp of the 5' end sequence were missing from the cDNA.

2.4. Localization of the Agr gene

The chromosomal location of the Agr gene was determined by in situ hybridization of Agr cDNA to polytene chromosomes from third instar larvae. The Agr cDNA was found to hybridize to the 12F region of the X chromosome (Fig. 4). No other hybridization signals were observed, suggesting that the Agr gene is unique. This was confirmed by Southern blotting of *Drosophila* genomic DNA and annealing of labeled Agr cDNA to the membrane-bound DNA at various hybridization stringencies (Doyle et al., 1990). Agr cDNA hybridized to a single 4 kb EcoRI-digested DNA fragment under conditions that promoted 0%, 20% or 30% mismatch between the probe and the target DNA (Fig. 5), indicating that the Agr gene is unique within the *Drosophila* genome, that there are no other closely related genes, and that the Agr gene is relatively compact and con-

tained within a 4 kb region. The uniqueness of the gene will simplify the genetic analysis of Agr protein function.

3. Conclusions

- (1) We have isolated a cDNA from *Drosophila melanogaster* that encodes a new member of the CAP protein family. This protein, termed Agr, is 256 amino acids in length and is most closely related to venom Ag5 of wasps. The Agr protein has a potential signal sequence and may be secreted to function as an extracellular enzyme.
- (2) Agr RNA is synthesized as a single 1000 nt transcript that is present during all stages of development. Highest levels of transcript are found during the adult stage. Adults of both sexes have similar levels of transcript.
- (3) The Agr transcript is synthesized from a single gene that is located within region 12F of the X chromosome. The 12F region is currently being mutagenized

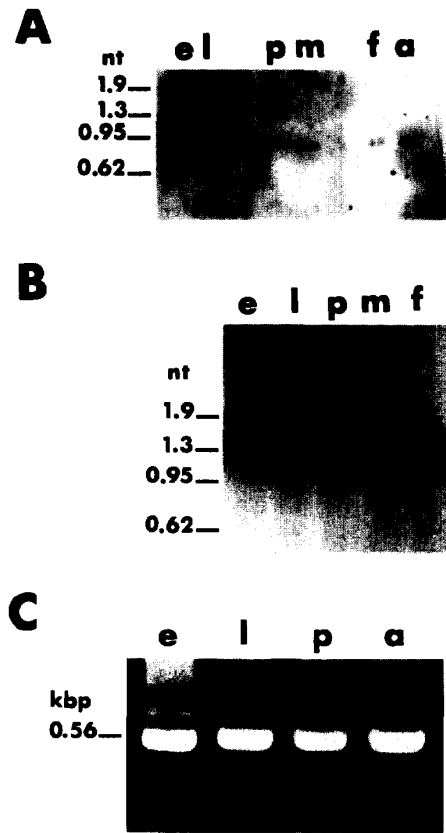


Fig. 3. Northern blot analysis of the *Agr* transcript from various developmental stages. Total RNA (15 μ g) was extracted (Chomczynski and Sacchi, 1987) from 0–16 h embryos (e), third instar wandering-stage larvae (l), pupae (p), adult males (m), adult females (f), and mixed sex adults (a), then size-separated by formaldehyde gel electrophoresis and transferred to nylon membrane. (A) Hybridization of antisense *Agr* RNA to total *Drosophila* RNA. The probe was digoxigenin (DIG)-labeled antisense *Agr* RNA. Hybridization was detected by the binding of an AP-linked anti-DIG antibody (Boehringer Mannheim) and by incubation with chemiluminescence substrates, followed by autoradiography. RNA size markers were run in lanes adjacent to the total RNA, transferred to membrane, and visualized with methylene blue. (B) Loading control. Identical amounts of total RNA from the indicated developmental stages were loaded and subjected to electrophoresis on the same gel as the samples in A. After transfer to membrane, the RNA from these lanes was stained with 0.2% methylene blue, destained in water, and photographed. The major RNAs in each lane were the 1730 nt a and the 2000 nt b fragments of 28S ribosomal RNA and the 2000 nt 18S rRNA (Beckingham, 1982). (C) PCR amplification of *Agr* cDNA from various developmental stages. *Agr* cDNA was synthesized from total RNA using an *Agr*-specific primer (complementary to nt 574–591 of the *Agr* cDNA sequence). The cDNA was then amplified by PCR using two *Agr*-specific internal primers (one primer had the complementary sequence indicated above; the second primer was identical to nt 22–39 of the *Agr* cDNA). Products of the PCR reaction were separated by gel electrophoresis and visualized by ethidium bromide staining. The predicted size of the amplified *Agr* product was 569 bp, which corresponded to the size of the major product of each reaction. The fragments in each lane could not be the result of amplification of genomic DNA, since the genomic DNA contains an intron sequence of approximately 100 bp (Kovalick, Karlo, and Schreiber, unpublished data) that would yield an

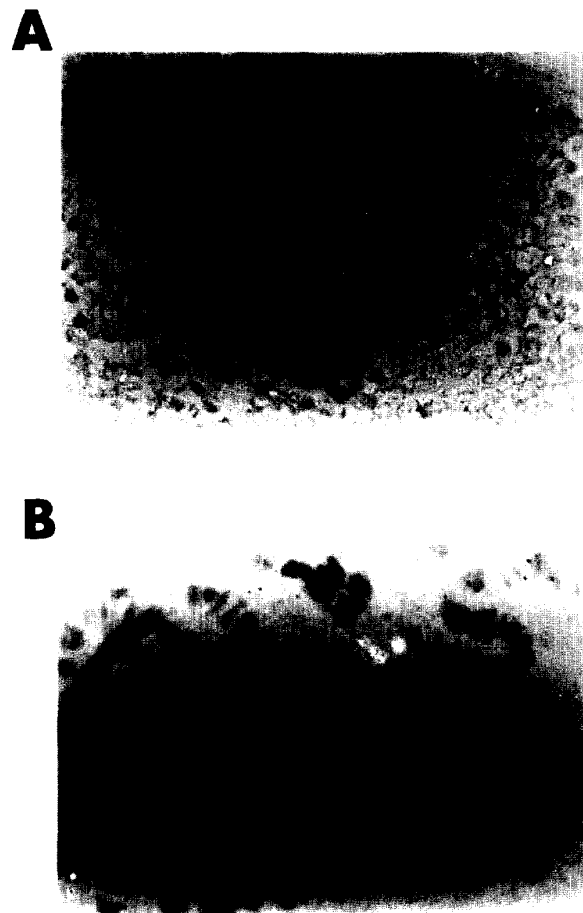


Fig. 4. In situ hybridization of *Agr* cDNA to polytene chromosomes. Salivary glands from third instar wandering-stage larvae were dissected and fixed in acetic acid, and polytene chromosomes from this tissue were prepared using the method of Atherton and Gall (1972). DIG-labeled *Agr* cDNA was hybridized to the chromosomes, followed by incubation with an AP-linked anti-DIG antibody and with the colorimetric substrates. Chromosomes were counterstained with Giemsa to visualize the banding pattern. Chromosomes were viewed and photographed under phase contrast microscopy. (A) Preparation showing a complete complement of chromosomes. The chromosomes are united at the chromocenter, located in the bottom center portion of the photograph. The site of *Agr* hybridization is indicated by a bar beside the X chromosome. Magnification 400 $\times$. (B) Detail showing the hybridization of *Agr* probe to region 12F. Major chromosomal regions near the site of hybridization are indicated in the photograph. Magnification 1000 $\times$.

amplified product of 670 bp. Since the PCR reactions were continued beyond the exponential phase of the amplification process, the amount of product obtained from each developmental stage does not reflect the initial amount of cDNA present in each reaction. The position of size markers is indicated at the side of each photograph.

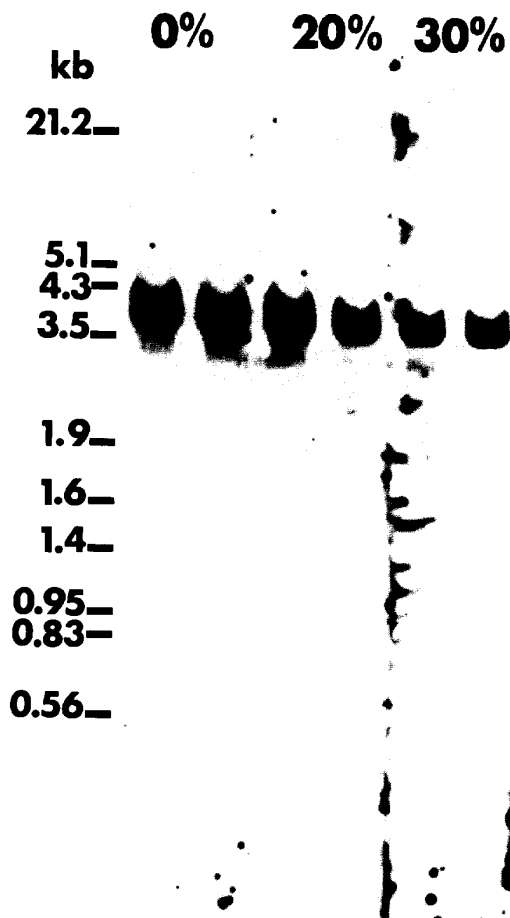


Fig. 5. Southern hybridization of *Agr* cDNA to genomic DNA at various stringencies. Genomic DNA was extracted from adult flies (McGinnis and Beckendorf, 1983), digested with the restriction enzyme *EcoRI*, separated by gel electrophoresis, and blotted onto nylon membrane (Southern, 1975). DIG-labeled *Agr* cDNA was hybridized to membrane-bound DNA under conditions that promoted either 0%, 20%, or 30% mismatch between the probe and the target DNA (Doyle et al., 1990). Hybridization of the probe was detected by incubation with an AP-linked anti-DIG antibody and chemiluminescence substrates, followed by autoradiography.

and screened for possible mutations within the *Agr* gene.

Acknowledgement

We would like to thank Dr. Richard Rayle (Miami University) for help with the sequence analysis; Mr. Matthew Mermer (Miami University) for advice and suggestions concerning the DNA sequencing protocol; and Drs. Jane Petschek, David Pennock, and Jack Vaughn (Miami University) for critical reading of the manuscript. These studies were partially supported by

Committee for Faculty Research grants from Miami University.

References

- Atherton, D. and Gall, J. (1972) Salivary gland squashes for in situ nucleic acid hybridization studies. *Drosophila Inf. Serv.* 49, 131–133.
- Beckingham, K. (1982) Insect rDNA. In: H. Busch and L. Rothblum (Eds.), *The Cell Nucleus*, Vol. 10, Part A, Academic Press, New York, pp. 205–269.
- Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P. and Tiver, K.K. (1986) Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoproteins secreted by the rat epididymis. *Eur. J. Biochem.* 161, 13–18.
- Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156–159.
- Cuasnicú, P.S., González Echeverría, F., Piazza, A.D., Camet, M.S. and Blaquier, J.A. (1984) Antibodies against epididymal glycoproteins block fertilizing ability in the rat. *J. Reprod. Fertil.* 72, 467–471.
- Devereux, J., Haerberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387–395.
- Doyle, K.E., Kovalick, G.E., Lee, E. and Beckingham, K. (1990) *Drosophila melanogaster* contains a single calmodulin gene: Further structure and expression studies. *J. Mol. Biol.* 213, 599–605.
- Fang, K.S.Y., Vitale, M., Fehlner, P. and King, T.P. (1988) cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5. *Proc. Natl. Acad. Sci. USA* 85, 895–899.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85, 8998–9002.
- Haendler, B., Kratzschmar, J., Theuring, F. and Schleuning, W.-D. (1993) Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/AEG) and the novel related CRISP-3 are expressed under androgen control in the mouse salivary gland. *Endocrinology* 133, 192–198.
- Hayashi, M., Fujimoto, S., Takano, H., Ushiki, T., Kazuhiro, A., Ishikura, H., Yoshida, M.C., Kirchhoff, C., Ishibashi, T. and Kasahara, M. (1996) Characterization of a human glycoprotein with a potential role in sperm-egg fusion: cDNA cloning, immunohistochemical localization and chromosomal assignment of the gene (AEG1). *Genomics* 32, 367–374.
- Hoffman, D.R. (1993) Allergens in Hymenoptera venom XXV: The amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. *J. Allergy Clin. Immunol.* 92, 707–716.
- Holton, T.A. and Graham, M.W. (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* 19, 1156.
- King, T.P., Sobotka, A.K., Alagon, A., Kochoumian, L. and Lichtenstein, L.M. (1978) Protein allergens of white-faced hornet, yellow hornet and yellow jacket venoms. *Biochemistry* 17, 5165–5174.
- King, T.P., Moran, D., Wang, D.F., Kochoumian, L. and Chait, B.T. (1990) Structural studies of a hornet venom allergen antigen 5, *Dol m V* and its sequence similarity with other proteins. *Protein Seq. Data Anal.* 3, 263–266.
- Kjeldsen, L., Cowland, J., Johnsen, A.H. and Borregaard, N. (1996) SGP28, a novel matrix glycoprotein in specific granules of human neutrophils with similarity to a human testis-specific gene product and to a rodent sperm-coating glycoprotein. *FEBS Lett.* 380, 246–250.
- Lu, G., Villalba, M., Coscia, M.R., Hoffman, D.R. and King, T.P. (1993) Sequence analysis and antigenic cross-reactivity of a venom

- allergen, antigen 5, from hornets, wasps and yellow jackets. *J. Immunol.* 150, 2823–2830.
- McGinnis, W. and Beckendorf, S. (1983) Association of a *Drosophila* transposable element of the roo family with chromosomal deletion breakpoints. *Nucleic Acids Res.* 11, 737–751.
- Miosga, T., Schaaff-Gerstenschlager, I., Chalwatzis, N., Baur, A., Boles, E., Fournier, C., Schmitt, S., Velten, C., Wilhelm, N. and Zimmerman, F.K. (1995) Sequence analysis of a 33.1 kb fragment from the left arm of *Saccharomyces cerevisiae* chromosome X, including putative proteins with leucine zippers, a fungal Zn(II)2-Cys6 binuclear cluster domain and a putative alpha 2-SCB-alpha 2 binding site. *Yeast* 11, 681–689.
- Morrisette, J., Kratzschmar, J., Haendler, B., el-Hayek, R., Mochca-Morales, J., Martin, B.M., Patel, J.R., Moss, R.L., Schleuning, W.D., Coronado, R. and Possani, L. (1995) Primary structure and properties of helothermine, a peptide toxin that blocks ryanodine receptors. *Biophys. J.* 68, 2280–2288.
- Murphy, E.V., Zhang, Y., Zhu, W. and Biggs, J. (1995) The human glioma pathogenesis-related protein is structurally related to plant pathogenesis-related proteins and its gene is expressed specifically in brain tumors. *Gene* 159, 131–135.
- Niderman, T., Genetet, I., Bruyère, T., Gees, R., Stintzi, A., Legrand, M., Fritig, B. and Mössinger, E. (1995) Pathogenesis-related PR-1 proteins are antifungal. *Plant Physiol.* 108, 17–27.
- Nishikawa, K., Kubota, Y. and Ooi, T. (1983) Classification of proteins into groups based on amino acid composition and other characters. I. Angular distribution. *J. Biochem.* 94, 981–995.
- Ohshima, M., Harada, N., Matsuoka, M. and Ohashi, Y. (1990) The nucleotide sequence of pathogenesis-related (PR) 1b protein gene of tobacco. *Nucleic Acids Res.* 18, 181.
- Pearson, W.R. and Lipman, D.I. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444–2448.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5468.
- Schuren, F.H.J., Asgeirsdottir, S.A., Kothe, E.M., Scheer, J.M.J. and Wessels, J.G.H. (1993) The *Sc7/Sc14* gene family of *Schizophyllum commune* codes for extracellular proteins specifically expressed during fruit-body formation. *J. Gen. Microbiol.* 139, 2083–2090.
- Sharp, P.M. and Lloyd, A.T. (1993) Codon usage. In: G. Maroni (Ed.), *An Atlas of Drosophila Genes*. Oxford University Press, New York, pp. 378–387.
- Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517.
- Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* 16, 8186.