Aspects of the regulation and function of *atonal*-like proneural genes in *Drosophila* neurogenesis

Eimear E. Holohan

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Edinburgh
January 2004

Declaration

I declare this thesis is my own work unless otherwise stated in the text, and that this work has not been submitted for any other degree or qualification.

Eimear Holohan

January 2004

Acknowledgements

First and foremost I would like to thank my supervisor Andrew Jarman for his help and guidance during the course of my PhD. Andy has been a great supervisor from whom I have learned a vast amount. I would like to thank him most sincerely for all the time and effort he has invested in me.

I would like to thank the members of the Jarman lab, Petra zur Lage, Lynn Powell, Davy Prentice, Emma Rawlins, Ivan Clark, Sam Maung and Sebastian Cachero. You've all been extremely generous with your time and expertise but a special thanks has to go to Petra for all her help ever since I arrived as a rotation student. Thanks must also go to other members of ICMB, especially the Davis lab, Ilan Davis, Hille Tekotte, Veronique Van de Boer, Renald Delanoue, Nina MacDougall, Isabelle Kos, Georgia Vendra, Alejandra Clark, Richard Parton, Ruth Bancewiec and Carine Meignin. I've really enjoyed being part of the Jarman/Davis lab and at times you have all provided a most welcome distraction! I would also like to acknowledge the assistance received from the cell sorting and Affymetrix technicians, Steve Le Moenic, Jan Vrana, Kevin Robertson, SCGT and Catriona Young, SHWFGF. Thanks of course to the Wellcome Trust for generously funding this work.

Thanks must also go to many others, especially to my parents for all they have done for me over the years, my sister Frances and my brother Daniel and to those people who have made my time in Edinburgh so enjoyable. Pride of place in that respect belongs to my flatmates throughout my PhD, David O'Regan and Meera Nair. Much more than just flatmates, I think David got it right when he described us as each other's Edinburgh family. You've both been absolute stars and the source of so much help, support and fun. Thanks also to my flatmates of recent months, Ed Galvin and Sebastian Cachero. A more specific thanks must go to Meera and Sebastian for help with the formatting and printing of this thesis. Another very special thanks has to go to Alejandra Clark, my friend both in the lab and out, whose friendship means a great deal to me. Others who have made my life significantly more enjoyable along the way include Mary Elliott, Henry Toulson, Aoife Kelleher, Ivor Waldron and members of the University of Edinburgh Hillwalking Club. Thanks most also go to my many Cork friends most of whom have managed at least one visit to Edinburgh over the years. Most notably, thanks to Linda Jackson, Jennifer O'Keeffe, Carina Burke, Áine Kennedy and Gillian Groeger.

So to all those who have helped in any way: Go n-éirí an bóthar libh. Or, for the non-Irish speakers amongst you: May the road rise up to meet you, may the wind be always at your back and may the sun shine warm upon your face.

Abstract

In *Drosophila,* all sense organ precursors require the function of a proneural gene for their specification. There are a number of proneural genes known, namely the *achaete-scute* complex, *atonal* and *amos,* each of which is responsible for a different type of sense organ. Proneural genes are expressed initially during development in equivalence groups known as proneural clusters and then in the sense organ precursor cells, which go on to form the sense organ. For *scute* and *atonal* these two phases of expression are under the control of distinct enhancer elements.

Unlike *scute* and *atonal* nothing is known of the *cis*-regulatory elements required for *amos* expression. Therefore one of the aims of my PhD was to identify the regulatory elements for *amos*. I was particularly interested in establishing if *amos* possesses autoregulatory enhancers and if this autoregulation is direct. As there are no known specific downstream targets of *amos,* identification of an autoregulatory enhancer provides a point of reference to compare other putative downstream target genes.

Despite the profound effects of proneural genes on neurogenesis, the study of differentially regulated downstream target genes has been severely curtailed by the scarcity of documented targets. A picture has yet to emerge of how the proneural factors function on a global, genomic level. Therefore in the second part of my PhD I focused on microarray analysis as a possible method for reliable identification of putative downstream target genes. I decided to pilot this approach by concentrating on the proneural gene *atonal*. My initial approach concentrated on whole embryos as

the source of RNA for microarray experiments. This pilot study has proved the feasibility of applying microarray analysis to the challenge of identifying proneural target genes, but also highlights the limitations of a whole embryo approach. A major limitation is that there is a high proportion of cells that are not relevant to sense organ precursor formation present in the sample. A clear improvement to the method of isolating RNA from whole embryos is to develop a system to obtain pure populations of sense organ precursor cells. As each proneural gene is responsible for a different subset of sense organs, the transcriptomes of these cells could be compared with each other by microarray analysis. I developed tools and methods for sense organ precursor cell sorting and concluded that this is both feasible and highly desirable for future progress in proneural target gene identification.

Contents

Abbreviations

ac – achaete
amos - absent md neurons and ofactory sensilla
aos- argos
AEL – after egg laying
APF – after puparium formation
AS-C- Achaete-Scute –Complex
ase – asense
ato – atonal
BDGP- Berkley *Drosophila* Genome Project
bHLH - basic-helix-loop-helix
Bearded – Brd
BPF– before puparium formation
BSA – bovine serum albumin
° C – degrees Celsius
cato – cousin of atonal
ch – chordotonal
CNS – central nervous system
CyO – Curly of Oster
da – daughterless
RNA – ribonucleic acid
DC – dorsocentral
RT – reverse transcriptase
Dl – Delta
DNA – deoxyribonucleic acid
ed – echinoid
EGFR-epidermal growth factor receptor
emc – extramacrochaete
EST – expressed sequence tag
es – external sensory
E(spl)C – Enhancer of split complex
FCO – femoral chordotonal organ
g – grams
GFP – green fluorescent protein
gcm – glial cells missing
klu – klumpfuss
L3/TSM – Vein L3/twin sensilla of the wing margin
lch – lateral chordotonal organ
l'sc – lethal of scute
loco – locomotion defects
lz – lozenge
mAb – monoclonal antibody
md – multiple dentritic
MF – morphogenetic furrow
mg – microgram
ml – millilitre
mirr - mirror
mM – millimolar
N – Notch
Ngn - neurogenin
nls – nuclear localisation signal
NP/PA – notoplural and post-alar
nvy - nervy
PCO – prothoraic chordotonal organ
PCR – polymerase chain reaction
phyl – phyllopod
PNC- proneural cluster
PNS – peripheral nervous system
pnt – pointed
PSC – presensillum cluster
RDA – representational difference analysis
rho - rhomboid
RGC – retinal ganglion cells
sc – scute
sca – scabrous
sens – senseless
SCGT – Scottish Centre for Genomic Technology
SHWFGF – Sir Henry Wellcome Functional Genomics Facility
SOP – sense organ precursor
Su(H) – Suppressor of Hairless
TCO – tibial chordotonal organ
TegCO – tegula chordotonal organ
ttk - tramtrack
vRCO – ventral radius chordotonal organ

1

Introduction

1.1 Neurogenesis in *Drosophila*

Neurons are formed by neurogenesis, a process that has much in common in different species. Neurogenesis in *Drosophila* initially involves the formation of neuroectoderm just dorsal to the ventral mesoderm. This region comprises ectodermal cells with the potential to form neural cells and epidermis. The neuroectoderm is subdivided into proneural clusters, from which one cell will emerge as the neuroblast and the remaining cells will become epidermal cells. About 30 neuroblasts are present per hemisegment and give rise to neurons, which will form the ventral nerve cord. The peripheral nervous system (PNS) is specified in a similar way, in that proneural clusters are selected in a precise spatial pattern, only one of the cells in the cluster adopts a neural fate and becomes a sensory organ precursor (SOP). Lateral inhibition mediated by the Notch signalling pathway, ensures that only one cell of the proneural cluster is fated to become the neural precursor cell (Simpson, 1997).

SOPs differentiate to form specific types of sense organs and all sensory neurons start axon and dendrite outgrowth shortly after their terminal division. The SOPs have their dendrites and cell bodies located in the periphery and their axons

projecting into the central nervous system (CNS). As *Drosophila* is a holometabolous insect it has both a larval PNS and an adult PNS. The larval PNS forms during embryogenesis and most larval neurons degenerate upon pupation. Adult sensory neurons form during the late third instar larval stage and pupal development. Many of the mechanisms and genes in *Drosophila* neurogenesis involved have counterparts in vertebrate neurogenesis.

Much work on neurogenesis has been done on the PNS of the fruit fly *Drosophila melanogaster*. This system has been described in great detail and there are many advantages to using it. The exact position and the times of emergence of the different types of sensory organ have been established. The sense organs consist of a relatively small number of cells and the lineage of these cells has by and large been determined (Brewster and Bodmer, 1996; Huang *et al.*, 1991; Jan and Jan, 1993). This means that the effects of any mutation can be investigated in detail. Other advantages include the knowledge of fly development and the available molecular genetic techniques.

The *Drosophila* PNS consists of five principal types of sensory organs: the external sensory organs (bristles), chordotonal organs (internal sensory organs, stretch receptors), multidentritic neurons, olfactory sense organs and photoreceptors (Jan and Jan, 1993). The pattern of sensory organs formed depends on where neural precursors are selected from the undifferentiated ectoderm during development (Ghysen and Dambly-Chaudiere, 1989). The proneural genes are central to this process.

1.1.1 Role played by proneural genes in neurogenesis

Proneural genes are key regulators of neurogenesis, facilitating the acquisition of specific subtype identities in addition to conferring a generic neuronal fate. Proneural genes are transcription factors of the basic-helix-loop-helix family and these bHLH genes are characterised by the possession of a basic domain to enable DNA binding to target genes and a helix-loop-helix for protein dimerisation.

There is a number of different proneural genes, the first discovered being the *achaete-scute* complex (*AS-C).* Subsequent work identified four genes as members of this complex, namely *achaete (ac), scute (sc), lethal of scute (l'sc)* and *asense (ase).* Loss of function and gain of function experiments have shown that *ac* and *sc* genes are required in the formation of most embryonic and adult external sensory organs as well as for a subset of CNS progenitors (Cubas *et al.*, 1991; Romani *et al.*, 1989; Skeath and Carroll, 1991). *l'sc* is expressed in the CNS primordium and is an important gene in the generation of neuroblasts (Jimenez and Campos-Ortega, 1990). The final member of the *AS-C* is *asense (ase). ase* is expressed later than the other members of the *AS-C,* is implicated in the correct differentiation of sense organs, and as such is considered to be a neural precursor gene. It is also the proneural gene for a set of wing margin bristles (Brand *et al.*, 1993; Jarman *et al.*, 1993a).

Another proneural gene *atonal (ato),* was discovered by Jarman *et al* in 1993. *ato* was isolated in a polymerase chain reaction to identify bHLH sequences similar to those found in the genes of the *ASC. ato* directs chordotonal organ (internal stretch receptor) formation in the embryo and adult (Jarman *et al.*, 1993b) and R8

photoreceptor formation (Jarman *et al.*, 1994). In addition *ato* is required for a subset of the olfactory sensilla, the sensilla coeloconica (Gupta and Rodrigues, 1997). In the CNS, *ato* does not appear to play a proneural role as loss of *ato* function causes defects in axonal branching and arborisation in larval brain (Hassan *et al.*, 2000). This implies that *ato* is required later in the CNS than the PNS.

amos, the most recently discovered proneural gene is responsible for the remaining subsets of olfactory sensilla, the sensilla basiconica and sensilla trichoidea and for two sensory neurons per segment in the embryo (Goulding *et al.*, 2000b; zur Lage *et al.*, 2003). *amos* was discovered by PCR amplification of *Drosophila* genomic DNA using degenerate primers. The primers were designed to the ends of the bHLH region that are conserved between Ato and its nearest vertebrate homologues. In addition to amplification of *ato,* two novel bHLH sequences were found. These were named *cato,* which is implicated in the differentiation of sense organs (Goulding *et al.*, 2000a) and *amos* (Goulding *et al.*, 2000b).

Neuronal bHLH proteins can be divided into two families based on the similarities between their bHLH domain. The proteins of the ASC share 70% homology in the bHLH domain with each other but only 42% with Ato. This compares with 88% identity between Ato and Amos and 64% identity between Ato and Cato. Therefore *ato, amos* and *cato* define a second family of proneural genes.

ASC, ato and *amos* between them appear to be responsible for specfication in the entire peripheral nervous system. However this is not the case in the CNS, where some neuroblasts do not require any of the known proneural genes (Jimenez and

Campos-Ortega 1990). Recently, new bHLH genes have been identified due to the sequencing of the entire *Drosophila* genome but none of these show the typical expression pattern of a proneural gene (Ledent and Vervoort, 2001; Moore *et al.*, 2000). Therefore if more proneural genes are to be found it is likely the structure will be different from those identified previously.

1.2 Sense organ precursor formation by the proneural genes

1.2.1 SOP formation by AS-C

Proneural genes are expressed in small patches (proneural cluster) of the developing ectodermal cells and give these cells the ability to develop into sensory organ precursor (SOP) cells. Cells within the proneural cluster compete with each other, so only one cell increases the expression of proneural genes and becomes the SOP cell. The neighbouring cells adopt the epidermal cell fate as the SOP cell prevents the remainder of the proneural cluster from adopting a neural cell fate. This is due to lateral inhibition which is mediated by the Notch signalling pathway (Simpson, 1997). In this process Notch signalling between the PNC cells activates the expression of transcription factors encoded for by the *Enhancer of split* complex which represses expression of the proneural genes (Jennings *et al.*, 1995). Delamination from the ectoderm of the selected precursor then occurs and it begins to express characteristic genes such as the neural precursor genes, among them *asense,* which is required for the determination and maintenance of neural precursor fate. The proneural genes are then shut off (Jarman and Jan, 1995).

1.2.2 SOP formation by *ato* and *amos*

This process in the case of the genes of the *ASC* ensures that only isolated external sense organ precursors are formed. The basic differences in sense organ precursors are in structure and sensory modality but differences in formation also exist. Large chordotonal arrays arise from clusters of neural precursors dependent on the proneural gene *ato*. Lateral inhibition mediated by Notch signalling still limits chordotonal precursor formation but clustering of embryonic chordotonal organs involves local recruitment mediated by the epidermal growth factor receptor (EGFR) (zur Lage *et al.*, 1997). Adult chordotonal organs contain far larger clusters of neurons and zur Lage and Jarman 1999, showed that this clustering results from the progressive accumulation of a large number of SOPs from a persistent proneural cluster. As in the case of the embryonic chordotonal organs, this is due to the opposing activities of Notch and EGFR signalling. While a two-step process is required for embryonic chordotonal organs, in adult chordotonal organ formation, EGFR acts in opposition to Notch by promoting continuous SOP recruitment (zur Lage and Jarman, 1999).

The development of *amos*-dependent olfactory sensilla is less well characterised but the preliminary steps seem similar to other sense organs. However *amos* expression is not switched off in the proneural cluster following SOP selection and the cells of an olfactory sensillum do not appear to arise solely by the division of the SOP cell (Ray and Rodrigues, 1995). Rather, the external sensillum structure is formed from the surrounding ectodermal cells that are recruited by the SOP.

1.3 Proneural genes as members of the bHLH family

The proneural genes are all members of the superfamily of the basic helix-loop-helix (bHLH) transcription factors. bHLH proteins function in a wide variety of tissues and developmental stages in a number of organisms to regulate precursor determination and lineage differentiation. In addition to neurogenesis, the bHLH proteins are involved in myogenesis and haematopoiesis (Porcher *et al.*, 1996) amongst others.

The bHLH domain is ~ 60 amino acids long and comprises a DNA-binding basic region (b) followed by two α-helices separated by a variable loop region (HLH) (Ledent and Vervoort, 2001). The bHLH domain is necessary and in some cases sufficient for the function of bHLH proteins (Chien *et al.*, 1996; Davis and Weintraub, 1992). The HLH region is required for dimerisation, allowing formation of either homo or heterodimers between different family members. The basic domain can be defined as a stretch of 12 amino acids beginning with the arginine residue common to all bHLH proteins (Hassan and Bellen, 2000) and it allows the bHLH proteins to bind to specific sequences in downstream target genes (Murre *et al.*, 1989). Since the bHLH motif was first defined by Murre *et al.* in 1989 by comparison between the proteins Daughterless, MyoD and Myc, many more bHLH proteins have been identified.

Atchley and Fitch in 1997 grouped the then known 122 bHLH sequences into four subgroups and named them A, B, C and D. Ledent and Vervoort (2001) searched extensively for all bHLH proteins in *C. elegans* and *Drosophila*

melanogaster, finding 35 and 56 respectively. A phylogenetic analysis of these genes in addition to >350 bHLH genes known, led the authors to alter and extend the earlier classification system of Atchley and Fitch 1997. The main differences were the addition of two new classes, E (includes *Enhancer of split)* and F, and, based on phylogenetic evidence the inclusion of class D within class A.

Class A includes the proneural proteins AS-C, Ato and Amos, their vertebrate homologues, as well as genes involved in muscle development such as *twist* and *MyoD.* Group D consisted of the HLH proteins that lack a basic domain and so are unable to bind DNA. This group includes the Id and Extramacrochaete proteins (Benezra *et al.*, 1990; Ellis *et al.*, 1990; Garrell and Modolell, 1990). In addition to its tissue specific constituents, class A also includes the ubiquitously expressed *daughterless (da)* and its vertebrate homologue E12.

1.4 How Proneural Proteins Work

The activity of the proneural genes requires that they bind to E-boxes (CANNTG) i.e. the *cis*-acting elements found in the promoters of downstream genes via the basic domain. (Hassan and Bellen, 2000). They must dimerise via the HLH domain with the bHLH product of the *da* gene in *Drosophila* which is expressed ubiquitously (Cronmiller and Cummings, 1993) and is required for every sense organ. Da forms dimers with AS-C (Cabrera and Alonso, 1991; Murre *et al.*, 1989) and Ato (Jarman *et al.*, 1993b) *in vitro* and there is genetic evidence that the same is true of Amos (Goulding *et al.*, 2000b). Loss of *da* results in the elimination of the entire PNS (Caudy *et al.*, 1988). Heterodimerisation is also an important mechanism

for mediating repression by HLH proteins. *Extramacrochaetae (emc)* encodes an HLH protein that forms heterodimers with Da and AS-C. As *emc* does not possess a basic domain and so cannot bind to DNA this results in an inactive dimer, which represses *AS-C* transcriptional activity (Van Doren *et al.*, 1991; Van Doren *et al.*, 1992). Similarly in vertebrates, tissue-specific bHLH proteins form dimers with ubiquitous bHLH proteins such as E12, a *da* homologue (Gradwohl *et al.*, 1996).

E-boxes have a core region of CANNTG and were originally found in the promoter region of the immunoglobulin heavy- chain (IgH) and the kappa light-chain (Ephrussi *et al.*, 1985). Since then E-boxes have been identified in the upstream regions of many of the genes involved in neurogenesis and other processes such as myogenesis (Lassar *et al.*, 1991). Class A bHLH proteins share a preference for the E box, CAGSTG, *in vitro*. In addition to regulating downstream target genes, the proneural proteins are also capable of autoregulation through binding to E-boxes in the upstream regions of the proneural genes themselves. Autoregulation is the term given to activation of a gene by its product. In the case of *sc* and *ato* there are well-characterised E-boxes present in enhancer elements that are responsible for accumulation of proneural protein in SOP cells (Culí and Modolell, 1998, zur Lage *et al* In prep).

1.4.1 Downstream target genes of the proneural proteins

Proneural genes are believed to regulate a common group of "generic" neural target genes for shared neural characteristics and for the lateral inhibition pathway. "Generic" target genes include *enhancer of split, Bearded, scabrous,* and *Delta*

(implicated in the lateral inhibition pathway), *senseless,* which is implicated in the maintenance of proneural gene expression, and the neural precursor gene *asense.*

In the PNS, loss-of-function mutations for each of the proneural genes, *AS-C, ato* and *amos,* result in the loss of a specific subset of sensory organs. Conversely gain-of-function experiments illustrate the ability of each proneural gene to induce particular types of sense organ when overexpressed (Goulding *et al.*, 2000b; Huang *et al.*, 2000b; Jarman *et al.*, 1993b; zur Lage *et al.*, 2003). Therefore genetic evidence suggests proneural genes are not only capable of conferring a generic neural cell fate on precursor cells but are also capable of determining specific subtype identity. It has been suggested that this specificity is due to selective regulation of different downstream targets (Jarman and Jan, 1995). This has proved difficult to assess directly due to the shortage of known differentially regulated target genes. There are however some candidates for *ato,* namely *rhomboid, TAKR86C* and *cato.* This has only been proven conclusively in the case of *TAKR86C* (Powell *et al.*, submitted).

1.4.2 How is the subtype specificity of proneural proteins regulated?

Protein-Protein Interactions

There are two possible alternatives for the mechanism whereby a proneural protein is able selectively to regulate downstream target genes. The first, widely held view, is that subtype specificity is due not to DNA binding differences between the bHLH transcription factors, but rather to their interaction with other protein cofactors

when bound to DNA (Brunet and Ghysen, 1999). This hypothesis is based on the similarity *in vitro* of the DNA binding properties of the proneural proteins (Jarman *et al.*, 1993b), and the conservation of the DNA-contacting residues of the basic domains in all proneural proteins. In 1996 Chien *et al.*, swapped the basic domains of *ato* and *sc* i.e. constructed chimaeric proteins. A Sc protein with an Ato basic domain was able to specify chordotonal organs in an *ato* mutant embryo, showing that the subtype specificity is in the basic domain. The basic domains of Ato and Sc differ by seven residues, and mutation of these residues suggests that most are required for chordotonal organ formation. A computer model predicts that none of these residues contact DNA directly which implies that it is likely Ato interacts with other protein cofactors when specifying neuronal subtype. The specificity of *amos* cannot lie in its basic domain as this region is identical to that of *ato*. Therefore specificity must be located elsewhere in the protein and may depend on cofactors.

Cofactors have been identified for a number of other bHLH proteins such as the myogenic bHLH family. Myogenic bHLH proteins interact with the MADS box transcription factor MEF-2 (Molkentin and Olson, 1996). In haematopoiesis the bHLH protein (SCL) specifies hematopoietic and vascular development by interacting with the LIM domain protein LMO (Porcher *et al.*, 1999).

There is a number of putative cofactors known for the neurogenic bHLH proteins. They include the Runt domain transcription factor encoded by the *lozenge* gene. Goulding *et al.*, (2000b) proposed *lz* as potential cofactor for *amos* based on their similarity of function, co-expression and strong genetic interaction. Strong *lz* mutants almost completely lack sensilla basiconica and show up to 50% reduction in

sensilla tricodea (Stocker *et al.*, 1993). In weaker *lz* alleles the major phenotype is transformation from basiconic to trichoid fate. *amos* expression in these alleles, patchy but spatially normal, suggests that higher levels of *lz* are required for SOPs to be fated to become basiconic while lower levels are sufficient for trichoid sensilla (Gupta *et al.*, 1998). Goulding *et al.*, (2000b) suggested that Lz could be a cofactor for the choice between basiconic and trichoid sensillum fate. It is noteworthy that *lz* may also function as a prepattern gene in the antennae (see chapter 2).

Another putative cofactor is the ETS domain transcription factor *pointed,* a component of the EGFR signalling pathway that plays a very specific role in the regulation of *ato* expression in the femoral chordotonal organ and in the embryo. Ato/Da/Pnt bind to sequences in the FCO enhancer and allow chordotonal recruitment to occur (zur Lage *et al.*, In prep). *ato* expression in other areas is under the control of different enhancers that do not require the action of *pnt* because recruitment does not occur. This means *pnt* is not generally required for *ato* activation of downstream target genes. Therefore *pnt* may be a cofactor for *ato* expression in the leg disc and embryo. The pattern of expression of *ato* and the greater diversity of *ato*-dependent SOPs (see chapter 2) together suggest *ato* may regulate downstream target genes in a more complex way that necessitates a greater involvement from regionally restricted cofactors.

From the evidence so far, it appears highly likely that cofactors exist which allow each proneural protein to specify distinct cell fates by facilitating interactions with the appropriate downstream target gene. Although the theoretical possibility existed that proneural genes were able to regulate differential downstream target

genes through differences in DNA binding, this had not been explored until very recently.

DNA Binding

Powell *et al.* (submitted) show that the common proneural target gene, *Bearded* (*Brd*), is regulated by Sc/Da and Ato/Da via distinct E_{Ato} and E_{Sc} sites. Singson *et al.* (1994) studied a 1.5kb fragment upstream of *Brd*. This fragment contains 10 E-boxes and supported *lacZ* expression in both the *AS-C* and *ato*-dependent PNCs. On the other hand, a smaller fragment of 180 bp containing only the *Brd-E1* site supported expression solely in the *AS-C*-dependent PNCs, thus indicating that an alternative E-box was responsible for expression in the *ato* dependent PNCs. Powell *et al.* showed a reporter construct of 1.3kb (i.e. lacking the 180 bp fragment) supported expression in the *ato*-dependent PNCs. This fragment contained 9 E-boxes and further constructs were made to subdivide this region. Expression was narrowed down to a region containing two E-boxes (E2 and E3). Mutation of E3 abolished reporter gene expression in the *ato*-dependent PNCs of the leg and eye disc. The sequence of this E-box is AA CACATG TT. This differs both from the consensus sequence for the *sc* dependent genes (G CAGSTG K) and from the E-box identified as being responsible for the autoregulation of *ato* in the SOP cell (A CAGGTG G). This is the first demonstration that DNA binding site differences underlie proneural specificity. Up until this point, it had been assumed that regulation of common target genes, by the different proneural proteins, occurred through the same E-box. This work showed that Ato and Sc appear to prefer and use different sites *in vivo*. This finding is substantiated by the observation that multimers of the different E box sites alone can confer highly specific patterns of expression on a reporter gene. These findings

establish that Sc and Ato-specific E-boxes have different regulatory properties *in vivo* and that differences in DNA binding sites can play a major role in proneural protein specificity.

1.5 Vertebrate Neurogenesis

Many vertebrate genes involved in neurogenesis are related to the *ASC* and *ato*. Vertebrate proneural-like genes can be divided into an *AS-C* like class containing amongst others *Mash1-2,* (Guillemot *et al.*, 1993; Johnson *et al.*, 1990) *Xash1* and *Xash3* (Turner and Weintraub, 1994) and those genes that show homology to *ato*. These include *Math1* and *Math5, NeuroD* (Lee, 1997)and *neurogenin* (Ma *et al.*, 1996). Just as in *Drosophila* where *ato* and *AS-C* code for different classes of peripheral sensory organs these vertebrate homologues are required for the development of different classes of vertebrate peripheral neurons.

However not all the vertebrate homologues have a clear proneural activity. In fact, proneural gene activity has only been established with certainty for *Mash1, Ngn1* and *Ngn2*. There is some evidence that the same is true of *Math1* and *Math5* but this remains to be proven. In loss of function experiments involving *Mash1* null mice, there is a loss of neuronal progenitors in the ventral telencephalon (Casarosa *et al.*, 1999) and the olfactory sensory epithelium (Guillemot *et al.*, 1993). *Ngn 1/2* are implicated in the formation of cranial sensory ganglia, spinal sensory ganglia and some of the ventral spinal cord neurons (Fode *et al.*, 1998; Ma *et al.*, 1998; Ma *et al.*, 1999; Scardigli *et al.*, 2001). As in *Drosophila* proneural gene loss-of-function experiments, Notch signalling is affected. In addition, gain-of-function experiments

in Xenopus have shown that *Ngn1/2* can ectopically activate neuronal differentiation and interact with Notch signalling (Ma *et al.*, 1996). These results suggest *Mash1* and *Ngn1/2* function as proneural genes in vertebrate neurogenesis.

The *ato* family member *Math1* is expressed in cerebellar granule cells, hair cells and touch receptor cells located throughout the skin (Ben-Arie *et al.*, 1997; Ben-Arie *et al.*, 2000; Bermingham *et al.*, 1999; Bermingham *et al.*, 2001). This indicates a functional similarity between *Math1* and *ato*. *ato* is responsible for the balance of the fly, through body wall and joint chordotonal organs and also for hearing, through the chordotonal organs which compose the Johnston organ, located in the second antennal segment. Ben-Arie *et al.* (2000) demonstrated that *Math1* can induce chordotonal organ formation in *Drosophila* and is capable of partially rescuing the *ato* mutant phenotype. In addition Wang *et al.* (2002), expressed *Math1* in *ato* mutant flies and *ato* in Math1 mutant mice. Unexpectedly, the two proteins were interchangeable. Although ectopic expression of both *Math1* in the chick neural tube (Gowan *et al.*, 2001) and of *Xath1* in *Xenopus* induces neuronal differentiation within ectodermal progenitors (Kim *et al.*, 1997), - indicating *ath1* possesses some proneural gene function - this still awaits confirmation. For *ath1* to be considered a proneural gene in the conventional sense, the progenitors of the lineages specified by *ath1* would have to be missing in loss-function-mutants. At this point in time there is no evidence that this is the case. While the same is true of *ath5* (the *ato* homologue expressed in the retinal ganglion cells (RGCs) of vertebrates), loss-of-function studies in the mouse and zebrafish show that most retinal ganglion cells are lost (Brown *et al.*, 2001; Kay *et al.*, 2001; Wang *et al.*,

2001). In addition, overexpression of *Xath5,* promotes the differentiation of RGCs in *Xenopus* (Kanekar *et al.*, 1997).

Therefore, currently there are three genes involved in vertebrate neurogenesis, namely *Mash1, Ngn1* and *Ngn2* that have a clearly defined proneural function. Furthermore, there is some evidence that the same is true of *Math1* and *Math5.* In addition, there is a number of other vertebrate homologues of the *Drosophila* proneural genes including *NeuroD.* Expression of *NeuroD* is activated after precursor selection and may play a similar role to that of *asense* and *cato* i.e. act as differentiation factors.

1.5.1 Target genes in vertebrate neurogenesis

The vertebrate homologues of the *Drosophila* proneural proteins also have the ability to interact with downstream target genes. *Ngn-3* is an islet- and neuron-specific proneural gene. Huang *et al.*, (2000a) showed that overexpression of *Ngn-3,* in islet-derived cell lines, induces a dose-dependent activation on the *NeuroD* promoter. The authors identified two E-box sites, (C CATATG G and A CAGATG G), and showed via mutational and gel shift analysis that a Ngn3- E47 dimer binds to these sites *in vitro.* Regulation of *NeuroD* by *Ngn* is analogous to regulation of *ase* by *sc* and *ato,* i.e. a proneural gene regulating a neural differentiation gene.

Roztocil *et al.* (1998) investigated the promoter region of the neuronal nicotinic acetylcholine receptor *β3.* This gene is expressed in the retinal neurons of the chick central nervous system and transient transfection assays identified a 75 bp region of the promoter, as being responsible for the neuron specific expression

pattern. This region contains an E-box of sequence A CAGCTG A. Mutation of this sequence abolishes promoter activity in retinal cells. Subsequent work showed that the *β3* subunit is under the specific control of the chick *ato* homologue ATH5 (Matter-Sadzinski *et al.*, 2001).

1.6 Aims of this study

The proneural genes play a profound role in neurogenesis. They are members of the bHLH family of transcription factors and not only confer generic neuronal characteristics but also specify subtype identity. This subtype specificity may occur through regulation of different downstream target genes and probably involves both DNA binding differences and protein interactions with subtype specific cofactors.

Unlike *sc* and *ato*, nothing is known about the *cis*-regulatory elements that govern *amos* expression. Therefore, one of the aims of my PhD was to establish if *amos* possesses autoregulatory enhancer elements and if this autoregulation was direct (chapter 2). The E-boxes of the autoregulatory enhancer elements of *sc* and *ato* contributed to establishing their respective consensus binding sites. Identification of the autoregulatory enhancer elements of *amos* (if any) and subsequent identification of the E-box responsible would shed light on the mechanism by which *amos* regulates target genes. It would also provide a point of reference to compare any future putative *amos* target genes.

All work carried out so far on the topic of proneural subtype specificity, has been of necessity piecemeal in approach and it does not give a clear picture of how the proneural factors function on a global, genomic level. In an attempt to remedy this, I focused on microarray analysis as a possible method for reliable identification of putative downstream target genes. I decided to pilot this approach by concentrating on the proneural gene *ato*. My initial strategy was to compare wild type embryos with mutant embryos and to compare wild type embryos with embryos in which *ato* was misexpressed throughout the neuroectoderm, i.e. wild type with both loss-of-function and gain-of-function. For gain of function I used *scabrous-Gal4* crossed with *UAS ato* to activate *ato* throughout the neuroectoderm. I subsequently decided to concentrate on gain of function experiments for reasons that will become clear in chapter three and compared *scabrous-Gal4* crossed with *UAS ato* with *scabrous-Gal4* crossed with *UAS nlsGFP*.

A limitation to this approach of taking whole embryos is the high proportion of cells in the sample that are not relevant to SOP formation. To overcome this problem I developed a system to allow isolation of pure populations of sense organ precursor cells (chapter four). These cells could then be used in microarray experiments to identify downstream target genes. Cell sorting as a means of obtaining pure populations of SOPs appears eminently feasible and certainly warrents further study.

2

Cis-regulatory elements required for the transcriptional control of *amos*

2.1 Aims of this Chapter

The main aim of this section of my PhD was to investigate the transcriptional regulation of the proneural gene *amos*. Unlike the other proneural genes, *sc, ac,* and *ato,* nothing is known about the cis-regulatory elements of *amos*. I identified enhancer elements by cloning fragments based on the sequence upstream of *amos* into a transformation vector containing a reporter gene, and by assessing the ability of different fragments to drive reporter gene expression in transgenic flies in a pattern of expression resembling *amos*. I then studied a number of fragments further. This involved a combination of genetic analysis and mutagenesis of specific sequences within these fragments to determine if they contributed to the reporter gene expression pattern and, therefore, to the expression of endogenous *amos*. In this chapter I intend to outline my contribution to elucidating how *amos* is transcriptionally controlled via *cis*-regulatory elements.

2.2 *amos* as a proneural gene

amos, the most recently discovered proneural gene, is responsible for two subsets of olfactory sensilla (Goulding *et al.*, 2000b; zur Lage *et al.*, 2003). The

adult fly possesses approximately 450 olfactory sensilla on the third antennal segment and 80 sensilla on the maxillary palp. The sensilla are divided into three main morphological subtypes, called sensilla basiconica, trichodea, and coeloconica. The coeloconica and some sensilla of the maxillary palp require *ato* (Gupta and Rodrigues, 1997). Goulding *et al.* (2000b) provided genetic evidence that the remaining subsets, the sensilla basiconica and trichodea are specified by *amos* and this was confirmed by the isolation of *amos* loss of function mutants (zur Lage *et al.*, 2003). *amos* and *ato* account for all olfactory sensilla found on the antennae.

In *amos* mutants, not only were olfactory sensilla lost but, surprisingly, ectopic bristles were formed on the third antennal segment. The appearance of ectopic bristles is thought to be due to derepression of *AS-C* and it suggests that, in addition to specifiying *amos*-dependent sense organs, *amos* also has a role to play in suppressing external sense organ fate. This is in agreement with an idea proposed by Jarman and Ahmed in 1998 whereby formation of bristles by the *AS-C* is the default state and *ato* must act to both inhibit bristle formation and impose chordotonal organ formation. The evidence for this lies in the observation that, while under certain well defined misexpression conditions *ato* is able to impose a chordotonal cell fate on existing bristle SOPs, the reverse is not true.

amos also has a role to play in the formation of the peripheral nervous system in the embryo. In *AS-C* and *ato* double mutants, only two sensory neurons remain per segment, the dbd and one of the dmd neurons (Jarman *et al.*, 1993). In *amos* mutants the dbd and one of the five dmd neurons are missing (zur Lage *et al.*, 2003).

This suggests that *AS-C, ato* and *amos* between them are responsible for the development of the entire embryonic peripheral nervous system.

2.3 Proneural gene expression

Proneural genes are key genes for neurogenesis and to a large extent the pattern of neurogenesis is decided by their pattern of expression. Not surprisingly, therefore, they are highly regulated and the study of their regulation is very important for understanding the patterning of neurogenesis. *AS-C* is expressed in a defined pattern mainly in the wing disc but also in the leg and eye-antennal disc. It is expressed initially in proneural clusters from which a single cell is selected by the process of lateral inhibition. Proneural gene expression is switched off in the proneural cluster and upregulated in the sense organ precursor cell (Cubas *et al.*, 1991; Skeath and Carroll, 1991). The SOP subsequently differentiates to become the macrochaete and microchaete of the notum and the campaniform sensilla of the wing blade.

AS-C is expressed in the embryo in the abdominal and thoracic segments, initially in clusters of ectodermal cells, and by stage 11 it is refined to expression in 11 SOPs and in an additional SOP present only in the first thoracic segment (Ruiz-Gomez and Ghysen, 1993; Skeath and Carroll, 1992). These give rise to external sense organs and to multiple dendritic neurons.

ato is expressed in all imaginal discs and is required for chordotonal organs (Jarman *et al.*, 1993), the photoreceptors of the compound eye (Jarman *et al.*, 1994) and for the sensilla coeloconica of the sacculus and the main antennal surface

(Jhaveri *et al.*, 2000; zur Lage *et al.*, 2003). Unlike sensory bristles, SOPs of chordotonal sense organs in the wing and leg disc are tightly clustered. In particular, the femoral chordotonal organ (FCO) contains some 70-80 scolopidia. SOPs continue to arise from a persistent proneural cluster due to the opposing effects of Notch and EGFR signalling (zur Lage and Jarman, 1999).

Embryonic expression of *ato* is also in proneural clusters in the abdominal and thoracic segments before being resolved to four or five cells in each segment. These are the precursors of the embryonic chordotonal organs (Jarman *et al.*, 1993). At this stage, chordotonal selection in the embryo appears very similar to external sense organ selection, i.e., expression in PNCs followed by Notch-Delta signalling to limit expression of the proneural gene to a single SOP. However, there is an additional step for chordotonal selection in abdominal segments whereby SOPs signal to adjacent ectodermal cells, via the epidermal growth factor receptor pathway, causing some of them to become recruited as additional chordotonal precursors (Okabe and Okano, 1997; zur Lage *et al.*, 1997).

2.3.1 *amos* expression during adult development

amos expression during adult development is restricted to the antennal discs from approximately puparium formation onwards. It is expressed in three semicircular bands in the third antennal segment which correspond to the sites from which a late wave of olfactory SOPs are selected. These SOPs are the precursors of the sensilla basiconica and sensilla trichodea (zur Lage *et al.*, 2003). As would be

expected for a proneural gene, *amos* appears to be expressed first in large proneural cluster domains and then in (*amos*-dependent) SOPs. Unusually however, PNC expression continues after the appearance of the SOPs. Therefore, *amos* proneural cluster expression does not appear to be totally inhibited by lateral inhibition following specification of the SOPs. Multiple SOPs are specified from each of the proneural domains over an extended period of time. These characteristics are atypical of proneural expression patterns and suggest that there will be unique features to the regulation of the *amos* gene.

2.3.2 *amos* expression during embryonic development

amos is also expressed in the embryo and promotes formation of a subset of multiple dendritic neurons (Huang *et al.*, 2000). At stage 10 of embryonic development, *amos* is expressed in clusters of cells in the head, the thoracic and abdominal segments. Expression in the head segments corresponds to the anlage of the olfactory sense organs of the larval antennomaxillary complex and it is switched off by stage 11. Expression in the thoracic and abdominal segments in stage 10 is in clusters of cells. This expression is restricted briefly to two cells and by stage 11 it is restricted to a single cell per segment. This suggests that lateral inhibition plays a role in singling out this cell and this is confirmed by the inability of *amos* expression to resolve to a single cell in Notch mutant embryos (Goulding *et al.*, 2000b). This conforms to the more orthodox expression pattern of the proneural genes.

In summary, *amos,* at least in the embryo, is expressed in the conventional proneural gene expression pattern, i.e., expression is initiated in a group of cells termed the proneural cluster and over time expression is resolved to a single sense organ precursor cell. However, expression of *amos* during adult development, although reminiscent of other proneural genes, is not strictly and solely in the conventional proneural expression pattern. Although *amos* expression is initiated in proneural domains, is not turned off in the surrounding cells following selection of an SOP.

It is perhaps worth noting that the conventional pattern of proneural gene expression (initially in the proneural cluster followed by the SOP) may be an oversimplification. It holds true for *AS-C, ato* and *amos* in the embryo (although there is an additional step for *ato* dependent precursors) however this is not the case for expression in the imaginal discs. In fact, *ac* and *sc* are the only proneural genes that follows this expression pattern at all times in discs. Although the expression of *ato* in the olfactory precursors of the antennal disc (Gupta and Rodrigues, 1997) follows convention, this is not the case in the leg disc. Here, at least in the FCO cluster, expression is not switched off in the proneural cluster following SOP selection. (zur Lage and Jarman, 1999). Instead, SOPs signal back to the proneural cluster to recruit more SOPs. Therefore, the fact that *amos* expression is not downregulated in the proneural cluster is not without precedence. There is no evidence to suggest that recruitment occurs in the same way in the *amos*-dependent SOPs of the antennal disc but it does remain a theoretical possibility however.

The widely differing modes of expression of *amos* indicate that more than one enhancer may be responsible for its expression patterns. I would predict the existence of distinct enhancers capable of transcriptionally regulating *amos* in a tissue specific manner.

2.4 Regulation of Proneural Gene Expression

Although, when compared to *sc* and *ato,* there are inherent differences in the mode of expression of *amos,* the basic pattern is the same. Similarly to *sc* and *ato, amos* is expressed in a proneural pattern, i.e., expression is initiated in the proneural clusters and later followed by expression in the sense organ precursor cells. For both *sc* and *ato*, these two phases of expression are driven by distinct enhancer elements. These enhancers are present in the surrounding regions of the genes and it is this transcriptional regulation that determines the position of the proneural clusters and, subsequently the position of the SOP. Having considered all available evidence I thought it highly likely that *amos* is regulated in a similar two-step way.

2.4.1 Enhancer Mediated Regulation of *atonal*

The *cis*-regulatory elements of *ato* have been studied by Yan Sun (Sun *et al.*, 1998). In this study, a comprehensive map of the enhancer elements around *ato* has been assembled. These enhancers were identified by the ability of fragments containing them to drive a reporter gene (*lac Z* in this case) in transgenic flies in patterns of expression resembling *ato*. A picture has emerged of *ato* being regulated by a series of enhancers both up and downstream of the open reading frame. A 3' enhancer directs expression in the proneural clusters of the chordotonal organs in the

embryo and larval leg and wing discs and also, partially, in the antennal imaginal discs. In addition, the 3' enhancer directs expression in a stripe anterior to the morphogenetic furrow in the eye. Sun *et al.*, (1998) proposed that successive modular enhancers located upstream of *ato* drive expression in the *ato*-dependent SOPs of the embryo, leg, wing and eye-antennal discs.

FIG 2.4.1

The expression patterns in the eye and the embryo directed by the 3' enhancer are significantly different from those directed by the 5' enhancer. The 3' enhancer directs expression anterior to the morphogenetic furrow and the 5' enhancer directs expression posterior to it. In embryos, the 3' enhancer drove expression in groups of cells surrounding the *ato* expressing SOP cell. These presumably correspond to the proneural clusters from which an *ato* dependent SOP was selected. The 5' enhancer appears to drive *lacZ* mainly, although not exclusively, in the embryonic SOP cells. It also drives expression in the Johnstons organ of the 2nd antennal segment at the third instar larval stage. Somewhat surprisingly, however, the expression patterns in the leg and wing discs, when driven by the 3' enhancer, closely resemble that of the 5' enhancer, i.e. the SOPs of the femoral chordotonal organ in the leg disc and the wing disc (Sun *et al.*, 1998).

2.4.2 Autoregulation of *atonal*

Autoregulation is the term given to activation or repression of a gene by its product. It is used to maintain and refine an expression pattern set up in response to a relatively short-lived signal. Activation may be direct or indirect. If direct, the gene product binds directly to enhancer elements and it stimulates expression of the gene. It is also possible that the protein does not itself bind the enhancer directly but

rather through a protein-protein interaction with a cofactor. The gene product could also stimulate expression indirectly either by enhancing the expression of an activator or by reducing the level of activity of a repressor. Autoregulation is a common regulatory mechanism and plays a role during cell fate specification in vertebrate tissues, notably muscle cells (Molkentin and Olson, 1996), in *C. elegans* neuronal cells - the LIM type homeodomain protein MEC-3, is required for the maintenance of its own expression (Xue *et al.*, 1993), and in other aspects of *Drosophila* development such as the regulation of the *fushi-tarazu* gene (Schier and Gehring, 1993). Autoregulation as applied to the proneural genes appears to allow a low level of expression in the proneural cluster to be followed by increased expression in the SOP cell.

Sun *et al.* (1998) examined the *lacZ* expression pattern of the 3' and 5' ato-lacZ reporter gene constructs in an ato^1 mutant background. The ato^1 mutation is a genetic null which allows Ato protein to be produced but the protein is non-functional (Jarman *et al.*, 1995). In leg discs, the *lacZ* expression from the 5' enhancer was completely abolished in the ato^1 mutant background whereas *lacZ* expression from the 3' enhancer was unchanged. In the eye disc, the 3' enhancer directed expression anterior to the morphogenetic furrow, whereas expression posterior to the MF, which is driven by the 5' enhancer, was eliminated. This suggests that expression of *ato* in the eye and leg discs is initiated via the 3' enhancer and that expression then becomes refined via the 5' enhancer by autoregulation. However the authors observed that *lacZ* expression is present in slightly altered patterns in the embryos and antennal discs when driven by the 5' enhancers in an *ato* mutant background. This suggests that either later expression in the 2nd antennal

segment and the embryo is not *ato* dependent or that residual atonal expression in the *ato* mutant is sufficient to direct reporter gene expression. However, following close examination of the expression pattern of *lacZ* driven by the 5' enhancer in the embryo, it is clear that there is a considerable amount of *lacZ* expression in the proneural cluster cells. This suggests that there may be both an SOP and a proneural cluster enhancer element present in this 5' region. In any case, the interpretation is complicated by the fact that SOPs fail to form in the *ato* mutant and so the loss of *lacZ* expression could simply reflect this failure rather than being due to autoregulation. More information might be gleaned from misexpression studies to determine if the reporter gene responds to an increase in *ato* expression. Genetic approaches indicate that autoregulation is occurring at least for the femoral chordotonal organ (FCO) of the leg disc but even here the evidence is not unassailable.

2.4.3 Autoregulation Direct or Indirect?

The autoregulation of *ato* may be direct or indirect. In general, this has not been explored at a molecular level except in one recent (unpublished) case of the leg femoral chordotonal (FCO) enhancer. The leg FCO enhancer has been further characterised by Petra zur Lage in the Jarman lab and consists of a 367bp fragment containing two E-boxes of sequence CAGGTG (unpublished observation). The presence of functional E-boxes (the binding sites for proneural proteins) is consistent with autoregulation being direct, i.e, Ato binds directly to an E-box in the upstream enhancer region and restricts expression to sense organ precursor cells. When cloned

into a transformation vector containing a reporter gene, the 367bp fragment supports GFP expression in the femoral chordotonal SOPs in the leg disc of the third instar larvae. When one of these E-boxes is mutated to CCATGG, GFP expression is abolished suggesting that the FCO enhancer is indeed responding to autoregulation and that this is mediated directly by Ato/Da binding to the E-box. DNA binding studies have confirmed the ability of Ato protein, when heterodimerised with the bHLH protein Daughterless, to bind to this E-Box in wild type but not mutated enhancers.

The existence of autoregulatory enhancers responsible for *ato* expression in other regions has not been explored in any detail. The presence of an autoregulatory enhancer for expression in the eye disc is suggested by the absence of *lacZ* expression posterior to the morphogenetic furrow, (driven by the 5' enhancer) in the ato^{1} mutant, (Sun *et al.*, 1998). This has not been explored further. In addition to expression in the eye disc and the FCO in the leg disc, *ato* is expressed in other regions, including the tibial chordotonal organ (TCO) in the leg disc. Expression in these regions is in clusters significantly smaller than that of the FCO. Sun *et al.*, 1998 showed that expression driven by the 5' enhancer in the FCO and the TCO in the ato^{1} mutant is abolished, indicating the existence of autoregulatory enhancers for both these chordotonal organs. The same enhancer may be responsible for autoregulated *ato* in the FCO and the TCO. *ato* is also expressed in the prothoraic chordotonal organ (PCO) of the proleg, in various other areas of the leg disc, in the ventral radius chordotonal organ (vRCO), the tegula CO (TegCO) of the wing disc, the Johnstons organ of the second antennal segment and in the precursors of the sensilla coeloconica of the third antennal segment. Given the mode of expression

patterns in these areas, there appears to be a reasonable chance that expression there is also subject to autoregulation, but this remains to be seen.

2.4.4 Enhancer Mediated Regulation of *AS-C*

The *ato* regulatory region is about 15kb in length. In contrast, that of the *AS-C* extends to about 90kb. Regulation of the *AS-C*, although more complex, has much in common with regulation of *ato*. Each proneural cluster is found in highly reproducible positions suggesting a significant degree of control of *ac* and *sc* transcription (Cubas *et al.*, 1991; Skeath and Carroll, 1991). SOPs then appear at specific positions within clusters and are defined by increased levels of *ac-sc* (Cubas *et al.*, 1991). Ruiz-Gomez and Modolell (1987) proposed that the complex patterns of *AS-C* expression are due to a battery of *cis*-acting regulatory regions found along the length of the complex. This proposal was based on the phenotypes of naturally occurring deletions of parts of the *AS-C* locus. The authors determined the phenotypes and molecular positions of the chromosomal breakpoints on the X chromosome. The mutants showed that the closer the breakpoint was to the *sc* gene, the more macrochaete were removed (Campuzano *et al.*, 1985). While allowing macrochaete associated with regulatory regions further downstream to form normally (Ruiz-Gomez and Modolell, 1987), deletions of regulatory regions upstream of *sc* led to the non-formation of all macrochaete corresponding to the putative regulatory regions within the deletion. A similar phenotypic analysis carried out with *ac* breakpoints gave similar results. It was therefore proposed that expression in the proneural clusters of the wing disc is accomplished by the action of prepattern

factors on highly site-specific *ac* and *sc* enhancer like elements found along most of the *AS-C*.

The proposal that *cis*-regulatory regions are responsible for the patterns of expression of *sc* is in line with work carried out by Leyns *et al.* (1989) who studied the role of the *AS-C* in the development of the campaniform sensilla on the wing blade. The authors showed genetically that the *sc* dependent campaniform sensilla appear to require control elements upstream of the *sc* gene itself. This is in contrast to *sc* dependent bristles, the majority of which depend on downstream control elements (Ruiz-Gomez and Modolell, 1987).

Thus the location of *ac, sc* enhancer elements had been proposed based on genetic evidence. Gomez-Skarmeta *et al.* (1995) then proved their existence at the molecular level. A number of different enhancers were identified for the proneural clusters found in the wing imaginal discs (Gomez-Skarmeta *et al.*, 1995). Each enhancer drives *ac* and *sc* in only one or a few proneural clusters (Ruiz-Gomez and Modolell, 1987; Skeath *et al.*, 1992). These enhancer elements are found either upstream, downstream or in between *ac* and *sc*. Thus the element responsible for expression in the notopleural and post-alar (NP/PA) proneural clusters is found downstream of both *sc* and *ac*. That responsible for expression in the presumptive vein L3 and at the twin sensilla of the wing margin (TSM) is found upstream of *sc* and downstream of *ac* and that responsible for expression in the dorsocentral (DC) region is upstream of both (Gomez-Skarmeta *et al.*, 1995). Similarly to *ato* 3' enhancers, these enhancer elements are thought to act independently of endogenous *ac* and *sc* and, as such, are not autoregulated.

The hypothesis is that specific prepattern factors interact with each enhancer to allow specific *ac* and *sc* expression in each proneural cluster. The DC enhancer is the best characterised at a molecular level. Pannier protein directly activates *ac* and *sc* by binding to this enhancer (Garcia-Garcia *et al.*, 1999). Ramain *et al.* (2000) showed subsequently that *Chip* (a ubiquitous nuclear protein) is required as a bridging factor to allow enhancer/promoter interactions. Other possible prepattern factors in the notum include the genes of the *iroquois* complex (Gomez-Skarmeta *et al.*, 1996; Leyns *et al.*, 1996). *araucan* is a member of the *iroquois* complex and ARA protein binds directly to the enhancer sequence responsible for *AS-C* expression in the vein L3 and TSM proneural cluster. In addition, there is a number of other genes that may be prepattern factors. The homeobox genes, *BarH1* and *BarH2,* are essential for the development of a subset of microchaete and the presutral macrochaete (Sato *et al.*, 1999). Since loss of *spalt* and *spalt-related,* which encode zinc finger transcription factors, removes the proneural cluster for the anterior macrochaete and suppresses the development of both these bristles (de Celis *et al.*, 1999), these genes appear to be responsible for formation of the anterior and posterior notopleural macrochaetae. There is no evidence, so far, that any of these proteins are capable of binding directly to *AS-C,* but it seems likely that this may be the case.

AS-C is also expressed in the CNS of *Drosophila.* CNS expression occurs at an earlier stage of development than PNS expression. By stage 8, *AS-C* is expressed in segmentally repeating clusters of cells which are arranged in columns along the ventral neuroectoderm (Ruiz-Gomez and Ghysen, 1993; Skeath and Carroll, 1992;

Skeath *et al.*, 1992). During stage 9, one cell, the neuroblast, retains proneural protein expression, while the other cells of the cluster lose this expression. The neuroblast loses proneural expression before it divides in late stage 9/early stage 10. *ac* and *sc* are not expressed in the neuroectoderm again until late stage 10. Skeath *et al.* (1992) utilised *lacZ* reporter gene constructs to identify the control regions present in the *cis*-regulatory region of the *AS-C* that are required for initial expression of *ac* and *sc*. Specific prepattern factors act on this regulatory region to confer positional control of proneural clusters in the CNS. The neuroectoderm is partitioned into domains dependent on the expression of combinations of prepattern genes. These prepattern genes are divided into those that are expressed along the anterior/posterior (A/P) axis such as the segment polarity genes *wingless, hedgehog, patched, gooseberry, engrailed* and *invected,* which are expressed in defined rows of cells (Bhat, 1999) and those that establish the dorsal/ventral (D/V) division of the neuroectoderm. These include *ventral nervous system defective, intermediate neuroblasts defective* and *muscle segment homeobox* (Gomez-Skarmeta *et al.*, 2003). The interplay of the anterior/posterior and the dorsoventral genes defines a prepattern, which in turn defines the position of the proneural clusters and confers a unique identity on neuroblasts.

2.4.5 Autoregulation of *scute*

Martinez and Modolell (1991) identified an enhancer in a 3.7kb region 5' from the *sc* gene that specifically drives accumulation of Sc protein in the sense organ precursor cells of the wing disc. Culí and Modolell, (1998) furthered characterised

this region and identified a 356bp minimal enhancer element, containing two E-boxes (E1 and E2). It also contains other conserved sequences that appear to be important in limiting *sc* expression to the SOP. This enhancer is dependent on endogenous *sc*, and mutation of E1 and to a lesser degree E2, reduced enhancer function. The authors also carried out gel retardation assays to examine the ability of Sc protein to bind to E1 and E2 boxes and found that a Sc/Da heterodimer bound both E boxes but with a higher affinity for the E1 box.

2.4.6 Autoregulation of *achaete*

In the region upstream of *ac* there are three E-boxes which are bound specifically *in vitro* by heterodimers of Ac and Da (also Da and Sc) (Van Doren *et al.*, 1991). This upstream region of the *ac* gene induced patterns of *lacZ* expression similar to those of endogenous *ac* in the proneural clusters and SOPs. Van Doren *et al.*, (1992) used a cotransfection assay to test the ability of the Da, Sc and Ac proteins to activate transcription from the *ac* promoter in *Drosophila* S2 cells. Transcription from the *ac* promoter increased 100 fold in the presence of expression vectors containing Da and Ac, or Da and Sc protein-coding sequences. In addition, mutation of the three E-boxes significantly reduces or removes reporter gene expression in the SOPs and PNCs (Martinez *et al.*, 1993; Van Doren *et al.*, 1992). These papers present strong evidence for the existence of an autoregulatory element present in the 0.9 kb upstream region of the *ac* gene, which mediates expression in both SOP and PNC cells.

2.5 Different methods of control for *AS-C* and *atonal*

The main site of expression of *AS-C* is the wing disc. It is expressed in more than 20 proneural clusters from which a pair of bristles will arise. Expression in each proneural cluster is dependent on an input from different prepattern factors and expression is upregulated in the SOPs via the same autoregulatory enhancer (Culi and Modollell, 1998). This is in contrast to *ato* regulation. *ato* is expressed in a more complicated pattern and is required for a wider range of sense organs i.e. chordotonal organs (Jarman *et al.*, 1993), photoreceptors of the compound eye (Jarman *et al.*, 1994), and for the sensilla coeloconica of the sacculus and the main antennal surface (Jhaveri *et al.*, 2000; zur Lage *et al.*, 2003). Unlike the case of the *AS-C,* different autoregulatory enhancers are required for *ato* expression in different regions (Sun *et al.*, 1998). This is likely to be due to different mechanisms of *ato* dependent SOP formation in different locations. For example, reiterative recruitment is needed to form chordotonal organs in the leg disc (zur Lage and Jarman, 1999), one round of recruitment is necessary for chordotonal organ formation in the embryo (Okabe and Okano, 1997; zur Lage *et al.*, 1997) and no recruitment requirement has been identified in the antenna or eye. This is in contrast to SOP formation by the *AS-C* complex, which is relatively uniform regardless of location.

2.6 Regulation and autoregulation of vertebrate homologues

There is a number of vertebrate homologues of the *Drosophila* proneural genes. The *AS-C* like family contains *Mash1-2, Xash1* and *Xash 3* and the *ato* like

family includes *Math1-3* and *Ngn1-2* (Hassan and Bellen, 2000). Similarly to *Drosophila,* vertebrate proneural genes are under the control of prepattern factors. Prepattern genes work to establish different progenitor domains in the spinal cord from which specific subtypes of neurons will arise. These genes include the vertebrate *Iroquois* genes (Gomez-Skarmeta and Modolell, 2002) and *Pax 6* (Scardigli *et al.*, 2003). In Xenopus, *Xiro* seems to control the expression of proneural genes such as *Xash3* and *Xngnr1,* as injection of mRNA promotes ectopic expansion of the neural plate, i.e. the domain of expression of these genes (Bellefroid *et al.*, 1998; Gomez-Skarmeta *et al.*, 1998). The prepattern gene *Pax-6* directly controls the spatial domain of activity of one of the *Ngn-2* enhancers (Scardigli *et al.*, 2003). Expression of *Ngn-2* in the neural tube is mediated by different enhancers which drive expression in distinct domains (Scardigli *et al.*, 2001). *Ngn-1* possesses regulatory elements that temporally and spatially control *Ngn1* expression in primary neurons of the zebrafish embryo (Blader *et al.*, 2003). In addition to prepattern factors, it is thought that diffusible molecules are also required for enhancer activity (Gomez-Skarmeta *et al.*, 2003).

Math1 expression is governed by multiple enhancers. Helms *et al.* (2000) identified two enhancers capable of driving *lacZ* reporter gene in the *Math1* expression domains. Unlike the enhancers identified for *Ngn1-2,* separable *Math1* enhancers for particular domains of expression were not found. The authors identified an E-box in one of the enhancers to which Math1 binds. *In vivo* expression of the *lacZ* reporter gene was abolished in *Math1* mutant embryos, indicating that vertebrate proneural genes are also capable of autoregulation (Helms *et al.*, 2000).

2.7 Investigating the enhancer elements of *amos*

Both *sc* and *ato* possess multiple enhancer elements in their regulatory regions. Moreover some of these govern expression in the proneural cluster cells and others in the sense organ precursor cells. The enhancer elements which stimulate expression of the proneural protein in the SOP cells may be autoregulatory, i.e. both *sc* and *ato* may be downstream targets of themselves. Unlike *sc* and *ato,* nothing is known of the *cis*-regulatory elements required for *amos* expression. A prediction would be that *amos* is also regulated by a number of discrete enhancer elements. Separate elements may be responsible for *amos* expression in different regions and these enhancers may respond to different prepattern factors and/or to autoregulation. Therefore, one of the aims of my PhD was to identify the regulatory elements for *amos*.

I am particularly interested in establishing if *amos* possesses autoregulatory enhancers and if this autoregulation is direct. This is because work carried out on the autoregulatory regions of *sc* and *ato* has not only contributed to the body of knowledge on control of proneural gene expression, but also to how proneural proteins selectively regulate different downstream target genes i.e. in this case they are downstream targets of themselves. This selective regulation takes place through the binding of the proneural proteins to E-boxes which are found in the enhancers of downstream genes. Recent work in the Jarman lab based on known downstream targets of *sc* and *ato* has identified a consensus sequence for the *ato* responsive enhancer elements which differs from the *sc* consensus. As there are no known specific downstream targets of *amos,* identification of an autoregulatory enhancer

will provide a point of reference to investigate how *amos* may specifically regulate its putative downstream target genes.

2.8 Identification of a region containing *amos cis*-regulatory elements

In an attempt to identify *cis*-regulatory elements for the *amos* gene, the upstream region of *amos* was analysed to identify the next closest gene. The intervening DNA of the upstream region was 3.6 kb in length. The downstream region was approximately 600 bp in length and was not investigated further. The upstream region had been previously cloned by Petra zur Lage into pBS and I cut out this fragment using restriction sites for Xba 1 and Asp 718. Following restriction digests, I ligated the entire fragment into the pPT-Gal4 vector. The Gal4 vector is a transformation vector that contains P element ends, a mini white selectable marker and a yeast Gal4 gene under the control of a minimal hsp70 promoter as an enhancer reporter (Sharma *et al.*, 2002). The resulting construct (*amos-3.6-Gal4)* was microinjected into syncytial blastoderm embryos and transformants obtained on the basis of eye colour. Two independent lines were established (*amosGal4 # 1* and *amosGal 4 # 2*) both with insertions on the second chromosome. The Gal 4 expression pattern of these lines was investigated by crossing either to *UAS GFP* or *UAS nlsGFP*. It was found that the 3.6 kb fragment supported GFP reporter gene expression in the pupal 3rd antennal segment in patterns reminiscent of endogenous *amos*. These patterns are described in detail in the following sections.

2.8.1 Expression of *amos-3.6-Gal4* in the antenna

White pupae were selected and aged for a set number of hours at 25° C. The pupae were dissected and the antennal discs fixed and mounted before being examined by confocal microscopy. GFP was observed in the third antennal segment. The expression pattern was further characterised by colabelling, initially using *amos* and *senseless (sens). sens* is present in all proneural derived SOPs and is thought to be a direct target of proneural genes (Nolo *et al.*, 2000). *amos* expression begins at puparium formation in 3 distinct semicircles and continues until 16 hours APF although the semicircles become indistinct by around 8 hours. SOP specification takes place in three waves (Ray and Rodrigues, 1995; zur Lage *et al.*, 2003). The first wave begins a few hours before puparium formation (BPF). SOPs are found in a semi circle at the outer edge of the third antennal segment. The second wave appears between 0 and 4 hours after puparium formation and gives three semi-circles of cells more centrally located. These two waves of SOP formation correspond to the sensilla coeloconica of the sacculus and the antennal surface and they are dependent on the proneural gene *ato*. In contrast, the third and final wave of SOP specification occurs at around 8 hours APF. The SOPs generated by the third wave accumulate between the rows of precursors generated during the second wave and are *amos* dependent. zur Lage *et al.*, 2003 showed that *amos* is responsible for the late wave of olfactory precursors. Colabelling with *amos* and *sens* suggested that the GFP coincides with the amos expressing SOPs only. This suggests that the 3.6kb fragment contains an *amos* SOP enhancer. In contrast, presumed *amos* expressing

proneural cluster cells do not express GFP, suggesting that PNC enhancer(s) exist outwith the 3.6kb region.

Perduring GFP expression, driven by the enhancer, can be observed after the downregulation of endogenous *amos*. This phenomenon can be exploited to follow the fates of the GFP (and therefore previously *amos* expressing) cells. This makes the enhancer a useful tool with which to establish both which sensilla, and which cells of each sensillum express *amos*. At 30 h APF, the different types of sensilla can be clearly distinguished morphologically. At this time, GFP is present in a number of the sensilla on the antennal surface. The sensilla morphology makes it clear that the amos-GFP expressing cells contribute only to the sensilla basiconica and sensilla tricodea. This confirms that it is the late forming SOPs that contribute to the sensilla basiconica and tricodea.

FIG 2.8.1/2 (A-C)

2.8.2 Expression of *amos 3.6 Gal4* during sensillum development

Following specification of an SOP, the next step is normally division of that SOP to form the sense organ (Jan and Jan 1993). However, in the case of antennal SOPs, at least for those belonging to the *ato* dependent 1st and 2nd wave, it appears that the SOP recruits neighbouring cells to form a pre-sensillum cluster (PSC) of two to three cells which then divide to form the different cells of the sensilla (Reddy *et al.*, 1997). These cells include the outer support cells (hair and socket cells), inner support cell (sheath cell) and 1-4 neurons. There is no direct evidence that cells derived from the *amos*-dependent SOPs form in a similar way but at 24 hrs each

basiconium and tricodeum consist of neuron, sheath and outer support cells. Specific markers can be used to recognise each of the different types of cells. At 24 h and beyond, the *amos* SOP enhancer drives GFP expression in most or all cells of the differentiating sensilla basiconica and trichodea. There is no GFP present in the cells of the sensilla coeloconica. The neurons are recognised by Elav expression, a RNA binding protein, and the sheath cell by expression of Pros, a homeodomain transcription factor (Sen *et al.*, 2003). The outer support cells are recognised by the expression of the homeobox gene Cut which is a selector gene for sense organs (Blochlinger *et al.*, 1991; Bodmer *et al.*, 1987).

The presence of GFP in the late PSC cells suggests that these cells derive from *amos*-expressing cells and that activation of the SOP enhancer is part of their specification process, although *amos* expression itself may not be long-lived in these cells.

FIG 2.8.1/2 (D, E)

2.8.3 Expression of *amos-3.6-Gal4* in embryos

Flies were allowed to lay overnight at 18° C, the embryos were collected and stained with antibodies to Amos and GFP. Widespread non-specific GFP expression was observed, probably due to chromatin position effects. This was the case with both lines analysed.

FIG 2.8.3

2.9 Expression of *amos-3.6-GFP*

The cloning of the upstream region of *amos* into a Gal4 vector resulted in the identification of an enhancer, whose expression recapitulates *amos* expression and was maintained in the SOP progeny of all cells of the sensilla basiconica and tricodea. This is in agreement with the *amos* loss of function phenotype, i.e. loss of sensilla basiconica and sensilla trichodea. The Gal4 construct is, however, not convenient for further study. The question as to whether or not this enhancer is autoregulatory has not been addressed. If it is indeed autoregulatory, it may respond to *UAS-amos*. In order to establish if this was the case, I cloned the 3.6kb fragment into a GFP vector called pH-Stinger and named the resulting construct *amos-3.6-GFP*. All further analyses were carried out on flies containing this construct.

2.9.1 Expression of *amos-3.6-GFP* construct in the antennae

The pH-Stinger vector is a *Drosophila* transformation vector that contains P element ends, a mini white selectable marker and nuclear GFP which is under the control of a minimal hsp70 promoter as an enhancer reporter (Barolo *et al.*, 2000). Using the restriction sites Xba1 and Asp718, the *amos* upstream region was excised from the *amosGal4* construct and cloned into the pH-Stinger vector. After microinjection, transformants were screened for on the basis of eye colour, and six stable lines established. *amos-3.6-GFP* lines 1, 5 and 6 have insertions on the second chromosome, line 2 has an insertion on the third and lines 3 and 4 have insertions on the X chromosome. Lines 1-4 were analysed in detail. Line 2 gave widespread non-specific GFP expression and so was not studied further. This may be due to

chromatin position effects, although the presence of Gypsy insulator sequences in the pH- Stinger construct should not allow this to happen.

Lines 1, 3 and 4 supported GFP expression in the *amos* dependent SOPs of the antennal disc (figure 2.9.1). Antennal discs were stained with α-amos and α-sens. *amos* is expressed in PNC cells as well as in SOPs whereas *sens* is expressed solely in SOP cells. If *amos-3.6-GFP* supported GFP expression solely in SOPs, all GFP expressing cells should also express *sens*. This, however, was not the case; some GFP expressing cells were negative for *sens* expression, but were positive for *amos* expression, indicating that these cells were PNC cells. This suggests that unlike *amos-3.6-Gal4, amos-3.6-GFP* drives GFP not only in the SOPs, but also in the PNC cells. This discrepancy between the *amos-3.6-GFP* construct in pH-Stinger and the same 3.6 kb fragment in *amosGal4* cannot be easily explained. One possibility is that there is a time difference in switching on the UAS GFP when driven by *amosGal4* as compared to the pH-Stinger construct and GFP is only switched on in *amosGal4* following specification of the SOPs.

It is not possible to state categorically which of the two constructs more faithfully reflects the wild type pattern of *amos* expression. This is because *amos* is expressed in two distinct phases i.e. expression in the PNC cells followed by expression in SOPs. The *amosGal4* construct supports GFP expression solely in the SOPs i.e. reflects the pattern of wild type *amos* expression only in the SOPs whereas the *amos-3.6-GFP* construct supports GFP expression both in the PNCs and the SOPs. As *amos-3.6-GFP* supports expression in both these regions I chose to concentrate on investigating the expression pattern of this construct. The work

outlined in the remainder of this chapter demonstrates that this construct recapitulates the expression pattern of *amos* in both SOPs and PNCs.

In any case, the 3.6kb fragment appears to contain elements that are responsible for SOP and PNC regulation in the antennal disc. It remains to be seen if these elements are separate or if both phases of expression are driven by the same enhancer element.

FIG 2.9.1

2.9.2 Expression of *amos-3.6-GFP* construct in the embryo

I investigated whether GFP was present in the *amos* expressing cells of the embryo. Lines 1 and 3 were used to determine the GFP expression pattern. I stained embryos of varying ages with α-amos and looked for colabelling of GFP. At stage 10 of embryonic development, *amos* is expressed in clusters of cells in the head, and thoracic and abdominal segments. There are five clusters of *amos* expressing cells in the head/thoracic regions which I have named 1-5 (figure 2.9.2), corresponding to the antennomaxillary cluster. In the head region, GFP driven by *amos-3.6-GFP* is seen overlapping with the largest cluster of *amos* staining (2). Immediately above this *amos* staining cluster (2), there is a cluster of *amos* expressing cells which does not express GFP at this stage (1). Another cluster of *amos* expressing cells in a more posterior location also expresses GFP (3). In embryo A (fig 2.9.2), GFP is not expressed in all cells of this cluster but in embryo B (slightly older) the GFP cluster is large and colabels more precisely with *amos*. Cluster (4) in embryo A is smaller

than that of embryo B. In embryo A, a cluster of *amos* expressing cells (5) is clearly seen in the posterior of the head region that does not express GFP. At stage 10, *amos* is expressed in clusters of cells in the thoracic and abdominal regions. In embryo A, one or two cells of the clusters also express GFP. In embryo B, more cells of each cluster are GFP expressing, ranging from two to approximately six.

By stage 11, *amos* expression in the head segments has been turned off. However, GFP driven by this construct perdures. Embryo C is an *amos-3.6-GFP* stage 11 embryo stained with an antibody to Amos. GFP staining can be seen in the head, which corresponds to previous *amos* expressing domains. Clusters 1-4 can be observed, by this time expressing GFP only. The uppermost cluster probably corresponds to cluster 1, which, in the previous images, did not express GFP, indicating that this is the last of the four head clusters to respond to *amos-3.6-GFP*. In addition, *amos* expression remains in cluster 5 suggesting, firstly, that *amos* expression in this region is longer lived, and, secondly, that in contrast to the other head clusters, at this time it is independent of *amos-3.6-GFP*.

At stage 11, *amos* is still expressed in groups of cells in most of the thoracic and abdominal segments. However, expression has disappeared in the most anterior thoracic and abdominal segments. Expression is present but is not yet been refined to a single cell in the more posterior abdominal segments: this suggests refinement to single cells is temporally separated in the different segments. Most of the GFP expression is located dorsal to the *amos* expressing cells with only a small overlap. This may imply that *amos* expression is switched off in a dorsal to ventral direction or that the GFP expressing cells migrate dorsally.

Embryo D is late stage 11 and *amos* expression is present only in cluster 5 in the head region. Unlike early stage 11, there may be an overlap of GFP within this cluster. *amos* expression in the abdominal region continues to be downregulated, with clear expression present only in some segments. By this time, *amos* expression has either been downregulated completely in all thoracic/abdominal segments or has been refined to a single cell. GFP overlaps with the remaining *amos* expression. In additional GFP expression is observed mainly dorsal to the *amos* expressing cell.

Again, I used GFP perdurance to establish the fate of these GFP expressing cells. I used 22C10 to mark all peripheral neurons in late embryos. There is a complex network of neurons in the embryo, but *amos* is responsible for only two of the multidendritic neurons, the dbd neuron and one of the five dmd neurons - marked in figure 2.9.2 E (Huang *et al.*, 2000). The dbd neuron has an associated glial cell. In *amos* mutants, these neurons are missing (zur Lage *et al.*, 2003).

GFP is observed in the head and in all thoracic and abdominal segments. In the abdominal segments, GFP expression is observed strongly in the more dorsally located *amos*-dependent dmd neuron in each abdominal segment. Less strong expression is observed in the dbd neuron and associated glial cell. The presence of GFP in the dbd and one of the dmd neurons confirms the ability of *amos-3.6-GFP* to drive expression in the *amos* dependent cells of the embryo. In addition, some ectodermal cells also express GFP. This expression is consistent with perdurance of GFP in some of the PNC cells which, due to lateral inhibition did not become SOPs.

GFP expression also perdures in the head region. Not all the GFP overlaps with 22C10 expression which indicates that here, also, some ectodermal or sense organ support cells express GFP. In addition, GFP expression is observed in the antennomaxillary complex. These are the presumed olfactory organs and so are suspected to require *amos* function, although this has not yet been shown.

In summary, GFP is found in the *amos*-dependent cells of the embryo but also in some ectodermal cells which were formerly part of the proneural cluster.

FIG 2.9.2

Judging from the expression of GFP driven by *amos-3.6-GFP* in the third antennal segment and in the embryo, it appears that the 3.6kb fragment upstream of *amos* may contain an enhancer element responsible for PNC expression in addition to SOP expression. If this is the case, it would be consistent with a model in which the proneural cluster enhancer will respond to as yet unknown prepattern factors and the SOP enhancer may depend on endogenous *amos* expression. The fragment was studied further with these two possibilities in mind.

2.10 E-box binding sites in the *amos* regulatory element

If *amos* is autoregulatory, and this autoregulation is direct as occurs with *ato* and *sc*, one would expect functional E-boxes to be present within the 3.6 kb region upstream of *amos*. Therefore, the 3.6kb fragment was scanned for E-box binding sites. A search using the very broad consensus sequence of CANNTG revealed 19

sites that matched this sequence (appendix D). Three of these matched the consensus sequence of *Enhancer of split* (a component of the Notch signalling pathway). This left 16 potential E-boxes through which *amos* might mediate autoregulation. In addition to the core six nucleotides, the flanking regions are also important for proneural protein binding and are considered part of the consensus. At the time this work was carried out, the known consensus sequence binding sites for Ato/Da was AA/T CAGGTG T/G. This was based on the binding site identified for the FCO enhancer (AA CAGGTG G, P. zur Lage, unpublished observation) and in the TAKR86C regulatory region (AT CAGGTG T, Rosay *et al.*, 1995). As *amos* is *ato*-like, the *ato* consensus sequence appeared to be a reasonable starting point to try and identify which (if any) of the 16 E-boxes might be Amos binding sites. Only one putative E-box of this sequence was found, AT CAGGTG A (differing only in the base pair immediately following the core). Moreover, this E-box is completely conserved in *Drosophila pseudoobscura* indicating that it may play an important role in the regulation of *amos*. Six out of the 19 E-boxes are conserved and the percent conservation in the upstream region between the two species is 50.1% suggesting any conservation may be relevant.

2.10.1 Mutation of E-box that fits *atonal* consensus

I decided to mutate this E-box to determine if it was required for activity of the *amos* enhancer. The E box was mutated, from CAGGTG to GGATCC. The bHLH domain makes major contacts with the CA and the TG, so the absence of these base pairs should prevent the interaction from occurring. Two independent lines were

established, 3.6 M1 (X chromosome) and 3.6M2 (2nd chromosome) and examined to determine if GFP expression was altered. GFP expression was still present, and following close examination, I could detect no discernible difference between the GFP observed in flies with the wild type enhancer compared with the mutated one. If multiple enhancers are present in the 3.6 kb fragment and only a subset were autoregulatory, then any change/loss of GFP due to autoregulation may be subtle. Since autoregulation is proposed to be most important in SOPs, I paid particular attention to the SOPs in the antennae (as marked by *sens* expression) in case there was a change in GFP expression in these cells. However I could observe no change. This appeared to hold true both for the 3rd antennal segment and the embryo. This suggests one of two things, either *amos*-3.6-GFP does not contain an autoregulatory element or, an alternative E box is responsible for any autoregulatory element that may be present. It is also noteworthy that the presence of a good proneural consensus sequence is not a perfect indicator of a true binding site.

FIG 2.10.1

2.11 Subdivision of *amos-3.6* region

From a number of observations it seems likely that multiple *cis*-regulatory elements are present in *amos*-3.6. In order to investigate better the possibility of an autoregulatory enhancer I needed to dissociate the enhancers present. The 3.6 kb fragment was split into three smaller fragments measuring approximately 1.6 kb, 1.0 kb and 1.0 kb - referred to as A, B, C. Each of these three fragments was cloned into the pH-Stinger vector, injected into embryos and stable lines were established. One line was established for A with an insertion on chromosome 2; two for B, B1 with an

insertion on the X chromosome, B2 with an insertion on the 2nd chromosome; and one for C (3rd chr). The GFP expression pattern (if any) driven by each fragment was examined.

FIG 2.11.1

2.11.1 *amos-A-GFP*

Fragment A (1.6kb) did not support GFP expression in either the antennal discs or in the embryo and, therefore, appears to contain at least no independent enhancer elements. However, I do have a concern regarding this construct in that unlike all the others, GFP expression is not observed in the salivary glands. Consequently, I cannot be sure that the line contains the *amos-1.6-GFP* construct. As I obtained only one line, I cannot confirm the observation that no GFP is present in either the 3rd antennal segment or in the embryo. It would be necessary to carry out PCR to amplify the 1.6 kb fragment and/or, GFP, to determine if this line contains the 1.6 kb fragment. Alternatively, another round of injections would provide more lines. Examination of these lines would then confirm or deny the absence of GFP.

FIG 2.11.2

2.11.2 *amos-B-GFP*

Fragment B did not support expression in the antennal disc. Fragment B contains the E-box that I had mutated, and the lack of GFP expression in the antennae confirms that this E-box is not important.

On the other hand, GFP expression was detectable in the embryo. The pattern observed in the head regions was, at stages 10 and 11, similar to that described earlier for the full length fragment. There may be more GFP expression in some of the head clusters in embryos containing *amos-B-GFP*, but this is difficult to establish accurately as any extra expression may be due to slight variations in the age of the embryos. The pattern of GFP in the head region, supports the hypothesis that an enhancer responsible for *amos* expression in the head region is present in fragment B.

However the pattern observed in the trunk region was different from that seen for the full-length fragment. GFP expression was present but did not overlap with *amos* in the thoracic and abdominal segments. Instead, expression was observed in the trunk region, in a segmental pattern that appeared to mimic expression in the head region. This expression is present at stage 10 in clusters of approximately 5 cells, located ventral to the *amos* expressing proneural clusters. The GFP expressing clusters increase in size and by late stage 11 range in size from about 8 cells to approximately 16. This ectopic expression pattern was observed in two independent lines.

amos-B-GFP late embryos were stained with 22C10 to mark all neurons and the expression pattern of the perduring GFP compared to that observed in flies containing *amos-3.6-GFP.* Unlike the case of *amos-3.6-GFP* there is no overlap of GFP expression with the dbd or dmd neurons. Instead, clusters of GFP expressing cells are located more laterally around the lch5 chordotonal organs. This association does not appear to be specific to the nervous system however.

FIG 2.11 3

2.11.3 *amos-C-GFP*

amos-C-GFP drove expression in the third antennal segment. GFP expression is present in the *amos* dependent SOPs and in some cells of the proneural cluster. There is no discernible difference between the location of GFP expression as driven by the *amos-3.6-GFP* fragment and that driven by *amos-C-GFP*. However, there may be a difference in the overall level of GFP expression. As I only obtained one *amos-C-GFP* line, it is difficult to say if this is a true reflection of the expression pattern or is line dependent.

In the embryo, there is no GFP expression at the times (stage 10/11) that *amos* is normally expressed, in either the head or the trunk region. However, as the neurons start to differentiate (as marked with 22C10) GFP expression is switched on in a subset of those cells marked by 22C10. GFP is present in two rows of internal cells. In a late embryo GFP is observed in a number of cells, some ectodermal and some associated with neurons, which do not appear however to be specifically associated with *amos*-dependent neurons.

FIG 2.11.4

In summary, *amos-B-GFP* contains an enhancer responsible for expression in the head region. GFP expression in the trunk region, is different from that observed with *amos-3.6-GFP,* and appears to unrelated to *amos*. This suggests that the enhancer responsible for amos expression in the trunk region lies outwith the 1.0kb

region being tested here. There is, however, ectopic trunk GFP expression, which appears to mimic GFP expression in the head region. This suggests that there is an inhibitory sequence outwith the 1.0kb fragment that normally restricts the activity of this enhancer to the head. The E-box tested previously is, apparently, not responsible for GFP expression in the head region of the embryo as expression does not appear to be altered in the mutated 3.6 kb fragment described previously.

amos-C-GFP, appears to contain an enhancer element responsible for *amos* expression in the third antennal segment.

None of these three fragments gave, in the trunk, an embryonic expression pattern that corresponded to the one obtained with the full-length fragment. This suggested that an enhancer element for this pattern was being split by the breakpoints of the subfragments. To test this, I made longer fragments containing AB and BC combined.

FIG 2.11.5

2.11.4 *Amos-AB-GFP*

Following microinjection I obtained five lines for *amos-AB-GFP*. Lines 1-3 had an insertion on the third chromosome, Line 4 had an insertion on the second and line 5 had an insertion on the X. Flies that have the fragment AB-GFP fusion lacked GFP expression in the antennal disc. This correlates with the lack of antennal GFP expression seen in flies with either of the two component fragments.

Interestingly, fragment AB produced an embryonic expression pattern similar to that obtained with the full-length fragment, even though its constituent fragments (A and B) do not support such a pattern. A alone does not support GFP expression in either the head or the trunk. B alone supports GFP expression in the head region and also ectopic expression in the trunk. This suggests that, as suspected, an *amos* embryonic trunk enhancer was indeed split by the separation of A and B. Furthermore, the ectopic trunk expression observed for fragment B was not observed for fragment AB. This supports the idea that B contains embryonic head sequences and A contains inhibitory sequences required to restrict activity of this embryonic head enhancer (assuming the flies do contain *amos-A -GFP,* see 2.11.1 for details).

FIG 2.11.6

2.11.5 *amos-BC-GFP*

I obtained one line for *amos-BC-GFP* (3rd chromosome). This line supported an antennal GFP expression pattern similar to that obtained with both the full length and fragment C, i.e., expression was observed in both the SOPs and the PNC cells.

In the embryo, *amos-BC-GFP* supports expression in the *amos* expressing cells in the head region. There is no GFP expression in the trunk. This is consistent with the lack of *amos* specific GFP in the trunk region driven by fragment B or C. Moreover the late appearance of non-specific GFP driven by fragment C, is not observed in *amos-BC-GFP.* This confirms that the late *amos-C-GFP* expression is likely to be an artefact of that construct.

FIG 2.11.7

2.11.6 Summary and Conclusions regarding the *amos* enhancer region

The results presented here show that the sequences required for *amos* expression in the embryo and antennal disc are separate. Hence, if *amos* is autoregulatory, different E-boxes would have to be responsible for expression in the antennae and embryonic *amos* dependent cells. Alternatively, these enhancers may solely respond to different prepattern factors required to regulate *amos* expression in the antennae and the embryo. In addition, as the head and trunk enhancers appear to be separate and more than one enhancer is responsible for *amos* expression in the embryo.

	A	B	C	AB	BC
Antenna	-	-	+	-	+
Head	-	+	-	+	+
Trunk	-	+ Ectopic	+ Ectopic	+	-

Table 2.11.1 Summary of expression pattern of different fragments

2.12 Further attempts to identify an amos-dependent E box

Mutation of the consensus CAGGTG core E-box in the 3.6kb fragment had no effect on enhancer activity. Subsequent analysis of subfragments revealed that this should not be wholly surprising because relatively little enhancer activity can be unequivocally assigned to region B, which contains this E-box. While region B does not drive GFP expression in the antennal disc, it does have a role to play in *amos* expression in the head region in the embryo. The E-box that I mutated was selected

as it best fitted the then known consensus sequence for Ato/Da binding site. As *amos* is *ato*-like it seemed reasonable that similar E-boxes would be responsible for control of expression. The chosen E-box, differed only from the then known consensus sequence in that the base pair immediately following the core was an A instead of T/G.

Since then, work carried out by other members of the Jarman lab has shown that both Sc and Ato are capable of regulating the same gene, *Bearded (Brd)* through different E-boxes (Powell *et al.*, submitted). The E-box through which Sc regulates *Brd* is CG CAGGTG T and the Ato specific E-box is AA CATGTG T (Singson *et al.*, 1994, Powell *et al.*, submitted). In addition to differing flanking regions, the core region of the *ato* specific E-box possesses a T in position 3 unlike the more commonly found G. This has two implications for putative *amos* dependent E-boxes:

1. the consensus sequence for *ato* has broadened considerably.
2. *ato* has the ability to regulate downstream targets through an E-box that differs from the *sc* consensus in the core region.

Perhaps *amos* specific E-boxes will differ more widely from the known *ato* consensus (although *amos* is *ato*-like). This means that many more of the 15 other E-boxes in the *amos* upstream region are potential Amos/Da binding sites. It seems reasonable to assume that the Da contacts will be similar to those already defined. Da has been shown to bind to the half site C/GTG G/T, however there are two precedents for proven Sc/Da sites that have a 3' A e.g. the Sc E2 site. There are two putative E-boxes that fulfil the apparent Da binding requirements. However, before

investigating any of the E-boxes, I sought to obtain direct experimental evidence for *amos* autoregulation. I did this by investigating the response of *amos* enhancer constructs to *amos* misexpression.

2.12.1 Misexpression analysis

Misexpression of the proneural genes in developing ectoderm leads to an increase in the number of SOPs formed and hence to an increase in sense organ formation (Rodriguez *et al.*, 1990, Jarman *et al.*,1993, Jarman and Ahmed 1998, Goulding *et al.*, 2000). *UAS-amos*, when driven by *109-68 Gal4,* a driver specific for proneural cluster and sense organ precursor cells (Jarman and Ahmed 1998) leads to an increase in neurogenesis in third instar larval discs. This can be visualised by α-Sens, and α-Amos. If the 3.6kb regulatory enhancer contains an autoregulatory component, then ectopic expression of *amos* may result in ectopic induction of *amos-3.6-GFP.*

I crossed virgin females containing *amos-3.6-GFP* line number 3 (X chr) to *109-68 Gal4* males. Male progeny from this cross were then mated with virgin females of genotype *UAS-amos.* I dissected third instar larvae discs, stained with antibodies to Sens, and Amos. Increased *amos* (from the UAS-construct) and *sens* expression and hence increased SOP commitment was observed in the eye-antennal, wing and leg discs of the third instar larvae. *amos* is not normally expressed in any of these tissues at this time in development. In the eye-antennal disc *amos* was expressed in the second and third antennal segment and also behind the furrow in the

eye. In the wing disc, expression was observed in all proneural clusters and in a stripe down the centre of the disc.

Examining GFP, I found that the *amos* enhancer did not respond to *amos* misexpression in most of these areas. However a response to misexpression was observed in the antennal disc. At 25° C, GFP expression was observed only in a subset of SOPs, (judged by *sens* staining), whereas at 29° C, GFP was present, not only in some of the ectopically induced SOPs, but also in some ectodermal cells. No GFP was observed in controls of this age. In order to determine if this GFP expression was specific to *amos*, the response to *UAS- ato* was also tested. *amos-3.6-GFP;109-68 Gal4* males were crossed to *UAS- ato* virgin females. This also gave GFP expression in the 3rd instar antennal disc. Expression however was even less widespread than that seen with *UAS- amos*, with a very small proportion of SOPs expressing GFP.

Therefore the *amos-3.6-GFP* responds to misexpression but is not completely specific for *UAS-amos*. Cross regulation of *amos* and *ato* has been observed before and, as such, the non-specificity of the *amos* is not surprising (S Maung, pers. comm.). These results indicate that *amos-3.6-GFP* does respond to *amos* misexpression, which is consistent with a degree of autoregulation. This expression occurs before *amos* is normally expressed and, consequently this expression is clearly temporally ectopic. However, as this expression is restricted to the antennal disc, it is not spatially ectopic and therefore the *amos* enhancer appears to be highly sensitive to location. GFP ectopic activation being so confined to the antennae it follows that autoregulation must be very contingent on the presence of other factors.

There is the possibility that an antennal-specific factor is required in addition to *amos*. This factor, if it does indeed exist, is clearly absent from leg and wing discs as judged by the inability of the *amos-3.6-GFP* construct to respond to ectopic expression of *amos* in any disc other than the antennal disc.

FIG 2.12.1

2.12.2 Search for New E-boxes

As there appears to be a degree of autoregulation at least in the antennal disc, associated with *amos-3.6-GFP* I again attempted to identify the E-box responsible. Of the remaining 15 putative E-boxes in the full-length fragment, four are present in *amos-C-GFP*. None of these four E-boxes fits the broadened consensus sequence for Ato. If Amos is interacting with one of these sites then it is appears to function quite differently to Ato. Of course, it is possible that, as more specific downstream targets of *ato* are discovered, this consensus will broaden even further and that the *amos* and *ato* consensus sequences will converge.

Of the four E-boxes two are conserved between *Drosophila melanogaster* and *Drosophila pseudoobscura* and one of these fits the apparent Da binding requirement. This box is TT CAAGTG A (C1). The final A is not conserved and is a G in *Drosophila pseudoobscura*. The other is GT CATTTG G (C4).

2.12.3 Mutating another E-box

amos-C-GFP gives GFP expression in the antennal disc. Therefore, I mutated the C1 E-box (conserved, fits Da binding requirement) to determine if expression

was altered/abolished. The E-box sequence is TT CAAGTG A and I mutated the core region to TT GGATCC A. I microinjected the construct and obtained one stable line with an insertion on the third chromosome which is homozygous lethal. I dissected pupal discs and stained with antibodies to Amos and Sens. When compared with *amos-C-GFP,* GFP expression was greatly reduced indicating that this E-box has a regulatory role in *amos* expression. If the GFP expression in the antennae were solely autoregulatory, I would expect expression to be abolished. If there were a contribution from separable SOP and PNC enhancers, expression should be abolished in the SOPs and remain in the PNC cells. However, GFP expression appeared to be uniformly reduced in all cells of the disc with weak expression being observed in some SOPs and PNC cells. This suggests that the enhancers responsible for expression in the PNCs and SOPs are not completely separate as appears to be the case for *sc,* but, rather, there is an interplay between the two. Also autoregulation may play a role in both PNC and SOP phases of expression.

FIG 2.12.2

2.13 Discussion

Proneural gene expression, initially in the PNC then followed by expression in the SOP, is driven by a series of enhancers. Each enhancer interacts with a specific combination of transcription factors. The position of a PNC at a particular location depends on a hierarchy of genes (Garcia-Garcia *et al.*, 1999; Rodriguez *et al.*, 1990). Positional information is specified by a group of transcription factors called prepattern genes. It is these prepattern genes which regulate the regional expression of the proneural genes. In addition, a number of enhancers appear to have an autoregulatory input, i.e. the proneural protein itself stimulates expression.

These two phases of proneural gene expression, at least in the cases of *sc* and *ato,* tend to be under the control of separate enhancers. It has been suggested that the enhancer responsible for expression in the PNC is under the control of prepattern factors while that of the SOP is autoregulatory. However, this may be an oversimplification. It remains entirely possible that, at least in some instances, the PNC enhancer is also subject to autoregulation. In addition, the SOP and PNC enhancers may not be as separable as has been thought.

As mentioned previously, Van Doren *et al.*, (1992) showed that the high level expression of an *ac-lacZ* reporter gene in PNC cells was abolished in an ac^{-} / sc^{-} background. A similar result was obtained when Ac/Da binding sites were mutated. This suggests that autoregulation has a role to play in *ac* expression in the PNC cells. Low level expression, probably under the control of unknown prepattern factors remained. Gomez-Skarmeta *et al.* (1995) identified enhancer regions which drive reporter gene expression in only one or a few PNC. These included the L3/TSM and the DC enhancers. In an ac^{-} / sc^{-} background, lacZ expression, driven by a *3.7sc-lacZ* gene (includes region responsible for expression in SOPs), is abolished in SOPs but remains in the L3/TSM PNC. Similarly, expression driven by the DC enhancer was observed in the DC PNC even in the absence of *ac* and *sc*. These results suggested that expression in the PNC cells was completely independent of the endogenous *sc* and *ac* genes, i.e. there is no input from autoregulation. However, this has subsequently been disproved in the case of the DC enhancer. Garcia-Garcia *et al.*, (1999) showed that *Pnr,* a GATA family transcription factor, directly activates *AS-C* by binding to the enhancer responsible for the expression of these genes in the dorsocentral proneural cluster. In addition, an autoregulatory step is mediated by

Chip, a ubiquitous nuclear protein. Chip bridges Pnr with the AS-C/ Da heterodimer which is bound to the so called SOP specific enhancer, leading to expression in the DC proneural cluster (Ramain *et al.*, 2000).

These findings have two implications for the study of proneural enhancers:

(i) Autoregulation has a part to play not only in *sc* SOP enhancers, but also in *sc* PNC enhancers.

(ii) *sc* SOP and PNC enhancers may not be as separable as previously thought.

The belief that *sc* SOP and PNC enhancers are completely separate arose from work carried out on the isolated enhancer of 356 bp. A reporter gene construct containing this fragment promotes expression solely in the *sc*-dependent SOPs (Culi and Modolell (1998). However, in its native state, it is also required for expression in the DC PNCs, indicating that this region plays a role regulating both SOP and PNC expression (Ramain *et al.*, 2000). This clearly illustrates the fact that an enhancer present in a reporter gene construct does not necessarily reflect what is happening *in vivo*. This is an additional complication when attempting to dissect the different contributions of autoregulation and prepattern factors.

Sun *et al.* (1998) studied the enhancer elements of *ato,* and suggested that a 3' enhancer drove expression in the PNCs and that a series of modular enhancers upstream of *ato* are responsible for expression in SOPs in specific areas. Again, however, at least in some instances, the separation of SOP and PNC enhancers may not be clear-cut. As mentioned previously, at least in the case of the embryo, there

may be both a PNC and an SOP enhancer present in the 5' region in addition to the PNC embryonic enhancer found downstream of *ato*.

Therefore, the previous delineation of SOP and PNC enhancers as being completely separate does not hold true. Similarly, the theory that autoregulation only has an input in regulating SOP enhancers appears to be an oversimplification.

2.13.1 Amos enhancer elements affecting *amos* antennal expression

I studied a 3.6 kb fragment upstream of *amos*. This fragment gave expression in the *amos*-dependent PNCs and SOPs of the antennal disc and embryo. It also responded to misexpression, albeit in a very specific fashion, i.e. GFP expression was only observed in the antennal disc. This suggests an antennal specific factor may be required in addition to *amos*. A similar observation has been made for the FCO enhancer, which only responds to misexpression when *UAS-pointed* is co-misexpressed with *UAS-ato* (zur Lage *et al.* in prep).

I attempted to dissect the enhancer elements further by subdividing the region. I found that *amos-C-GFP* gave an expression pattern in the antennae similar to that obtained with the full- length fragment. Of the four E-boxes present in this region, two are conserved, one of which fits the apparent Da binding requirements. Therefore, I mutated this E-box. This site is conserved in *Drosophila pseudoobscura,* (apart from the final A) and it clearly plays a part in *amos* regulation as judged by an overall reduction in GFP expression in the antennae with the mutated construct. As expression is not reduced specifically in SOPs, but rather in both SOPs

and PNCs, the enhancer does not appear to be solely an SOP one. This suggests that if Amos/Da is binding to this site, autoregulation may play a role in both PNC and SOP phases of expression. If autoregulation is occurring, the *amos-C-GFP* fragment should respond to misexpression. Somewhat surprisingly, this proved not to be the case (data not shown). However, as I only have one line, this may be an artifact of that particular line. In order to establish with certainty if *amos-C-GFP* is capable of responding to misexpression, I would need to reinject to obtain more lines.

In addition, the full-length fragment responds to misexpression only in the antennae, indicating the existence of an antennal specific factor. If this is the case, and if like Pointed (implicated in the response to misexpression by the *ato*-dependent FCO enhancer), this factor binds to the enhancer, it is possible that the binding site for this additional factor is present outwith the *amos-C-GFP* region. This would also provide a possible explanation for the apparent lower level of GFP expression in *amos-C-GFP* when compared with the level of expression in the full-length fragment. This apparent lower level of GFP expression remains to be confirmed by the analysis of more lines however.

2.13.2 Mutant Background

Another method of determining if the *amos* enhancer is autoregulatory is to test its response to *amos* mutation. The rationale here is that if no functional Amos protein is being produced, then there will be no activation of an *amos* autoregulatory enhancer. In contrast, upstream prepattern factors will not be affected. The complication with this, however, is that these mutations always cause loss of SOPs.

Therefore, expression of GFP in SOPs will of course be abolished regardless of whether such expression results from an autoregulatory enhancer or not. This complication is often overlooked. As low level GFP expression remains both in the PNCs and the SOPs when the putative Amos/Da binding site is mutated, this suggests the site is important in both PNC and SOP expression; it also suggests that autoregulation may be involved in mediating expression in the PNCs as well as the SOPs. If this is the case, high-level expression of GFP driven by the enhancer in the PNCs should be abolished in *amos* mutant background. If autoregulation does not have a role to play in the PNCs, expression in these regions should be unchanged. Expression of the constructs in *amos* mutant backgrounds will help to establish if the PNC enhancer has an autoregulatory input.

2.13.3 Does *amos* bind directly to the putative Amos/Da binding site?

Previous work suggests proneural proteins bind directly to E-boxes. The evidence for this is based mainly on *in vitro* binding studies where proneural proteins bind to wild type but not to mutated enhancers (Culi and Modolell, 1998, Van Doren *et al.*, 1992, zur Lage *et al.*, in prep). In actual fact, the direct binding of any proneural protein to an E-box *in vivo,* has not been proven. Although *in vitro* binding occurs and mutation of the E-boxes involved abolishes reporter gene expression, work up to now cannot rule out the possibility that, *in vivo,* a protein or protein complex not involving the proneural protein is responsible for the E-box dependent activation of proneural gene transcription. However such a hypothetical activator, must

(i) be dependent for its expression or activity on proneural gene function, and

(ii) require the integrity of the core nucleotides of the E-box sites to which proneural protein/Da complexes bind *in vitro*.

While this suggests that the most likely explanation is that proneural/Da binding is indeed direct, this has not been established with certainty.

Therefore, although the mutated site in this case is clearly important as judged by the decrease in GFP expression, I cannot be certain that it is an Amos/Da binding site. *In vitro* DNA binding is unlikely to provide much more information as proneural protein/Da heterodimers bind non-selectively to CANNTG. It would be necessary to carry out ChIP to establish with certainty if Amos/Da bind directly to this site *in vivo*. This is also necessary for the binding sites identified for the other proneural proteins, as direct binding has not been conclusively proven.

However, as mentioned previously, multimerised 20 base pair regions including the E-box and the flanking sequence of *sc,* recapitulate *sc* specific SOP expression in the embryo. The same is true of the E box implicated in *ato* dependent expression in the SOPs of the FCO, i.e. a multimer of that E-box and its flanking regions drives GFP expression in the *ato* dependent SOPs of the embryo. It appears likely that the same may be true of an E-box to which Amos binds and therefore a multimer of the putative Amos/Da binding site may drive GFP expression in the *amos*-dependent SOPs of the embryo. This would be further evidence that the putative Amos/Da binding site is involved in *amos* regulation.

The evidence so far suggests that the site mutated is important in mediating *amos* expression in the PNCs and SOPs of the antennal disc. This regulation may occur through Amos binding or through the binding of some other as yet unknown factor. However, mutation of this site does not result in the complete absence of GFP. Low-level GFP expression remains. This indicates that there may be other important sites in the enhancer that are responsible for this expression. There are three other E-boxes in this fragment, in particular C4 (conserved between *Drosophila melanogaster* and *Drosophila pseudoobscura),* which may be responsible for the remaining GFP expression. An alternative possibility is that the low level GFP expression may be the result of the action of a prepattern factor such as *lozenge.*

2.13.4 *lozenge* as a Candidate Prepattern Gene for amos expression in the antennae

Goulding *et al.* (2000b) provided evidence that *amos* expression prior to SOP formation is regulated by *lozenge (lz). lz* possesses a DNA binding domain similar to that of the Acute myeloid leukaemia-1/Runt transcription factors and it plays a role in the development of the eye, antennae and tarsal claw. It is expressed in the antennal disc in a manner strongly resembling the *amos* expression pattern albeit in an earlier time period (Goulding *et al.*, 2000b). Strong *lz* mutants almost completely lack sensilla basiconica and show up to 50% reduction in sensilla tricodea (Stocker *et al.*, 1993). Although some alleles have been shown to affect the morphology of these sensilla, there is no change in the number of coeloconica (Gupta and Rodrigues, 1995). In strong *lz* mutants, *amos* RNA was missing from the middle of all three antennal bands (Goulding *et al* 2000b). Sensilla basiconica are fated to arise from the middle of band 3 (Gupta *et al.*, 1998) and it appears likely that the middle regions

of bands 1 and 2 give rise to the sensilla trichodea missing in strong *lz* mutants. In weaker *lz* alleles, the major phenotype is transformation from basiconic to trichoid fate. *amos* expression in these alleles is patchy but spatially normal, suggesting that, while lower levels are sufficient for trichoid sensilla, higher levels of *lz* are required for SOPs to be fated to become basiconic (Gupta *et al.*,1998). Activation of ubiquitous *lz* expression using a hs-*lz* construct (Gupta *et al.*, 1998) led to a strong expansion of *amos* expression in the antennae and also resulted in ectopic amos expression in the leg and wing discs (Goulding *et al.*,2000b). These results are highly suggestive of *lz* fulfilling a prepattern role in the antennae. If this is the case, the low level GFP expression observed in the antennae of flies containing the mutated version of *amos-C-GFP,* may be under the control of *lz*. This control may be direct or indirect but the presence of three *lz* binding sites, [(RACCRCA -(Flores *et al.*, 2000), (RRCGCA - (Xu *et al.*, 2000)] within the *amos-C-GFP* fragment suggests that *amos* may be under the direct control of *lz*. Mutation of these sites to determine if GFP expression is abolished, misexpression analysis to determine the effect of overexpression of *lz* on the enhancer and a study of the enhancer in a *lz* mutant background could be carried out to determine if *lz* is capable of regulating this enhancer.

2.13.5 Summary regarding antennal expression

The *amos-C-GFP* fragment drove GFP expression in the *amos*-dependent PNCs and SOPs of the antennal disc. Mutation of one of the two conserved E-boxes led to an overall reduction in the expression levels of GFP i.e. both PNCs and SOPs. Unlike *sc* and *ato,* there is no evidence to suggest the existence of an SOP only enhancer for *amos*. Rather, the two phases of expression, i.e. expression initially in

the PNC followed by expression in the SOPs, seem tightly linked. Although the hypothesis is that Amos/Da binds directly to this E-box, there is no direct evidence that this is the case. Further misexpression and mutant analysis should help resolve this point.

2.13.6 Embryonic expression

The full length fragment gives expression in the *amos* dependent cells of the head and trunk regions of the embryo. There appear to be different enhancer elements for the head and trunk. Fragment AB, - the only fragment to support a GFP expression pattern in the embryo similar to that of the full length fragment, - contains 15 putative E boxes of which four are conserved between *Drosophila melanogaster* and *Drosophila pseudobscura*. Therefore the enhancer responsible for expression in the trunk lies within this region. In addition, fragment BC supports GFP expression in the *amos* expressing cells of the head region. This suggests that the enhancer responsible for *amos* expression in the head is found in region B. However the GFP expression driven by fragment BC in the head is very much less than that driven by AB. It is possible a region of A is necessary to provide full expression or this may be simply due to variations between lines.

Region B contains six E-boxes, one of which on further testing, does not appear to contribute to expression in the head region. Of the remaining five, two are conserved in *Drosophila pseudoobscura* and as such appear to be the most likely candidates. Future work will involve transient injection assays in embryos to determine the E-boxes responsible for each of the different areas of embryonic *amos*

expression. Future work will also involve misexpression in the embryo to determine if there is an autoregulatory component. It is also possible that some GFP expression may be under the control of a prepattern factor. One possible candidate is *mirror*.

2.13.7 *mirror* as a candidate prepattern factor of amos in the embryo

mirror (*mirr*) is a member of the *Iroquois* complex and was originally identified as being involved in eye development (McNeill *et al.*, 1997). Judging by its embryonic expression pattern, *mirr* is a possible candidate to regulate the *amos* specific enhancer. In addition, the *amos* dependent dbd neuron is abolished in a *mirror* mutant embryo. This suggests that future work should be aimed at looking for *mirr* regulation of *amos*. As a preliminary, I looked at *mirr* consensus binding sites in *amos*-3.6 in *Drosophila melanogaster* and *Drosophila pseudobscura*. The *mirr* consensus binding site is ACANNTGT (H. McNeill, pers. comm.) and I located three of these sites. However, none of these sites is conserved between the two species. This reduces the likelihood of their being responsible for *mirr* binding but this possibility could be tested, by mutating the putative *mirr* binding sites, to determine if proneural cluster GFP expression is abolished. Misexpression analysis could also be used to examine the consequences of ectopic expression of *mirr* on *amos* expression and on GFP expression driven by the 3.6kb enhancer. Expression of the enhancer in a *mirr* mutant background to determine if expression is abolished would also be informative.

2.14 Conclusions

This work has demonstrated that the enhancers responsible for *amos*-dependent expression in the antennae and embryo are separate. Not only that, but distinct enhancers appear to be responsible for expression in the head and trunk regions of the embryo. In addition, I identified a putative Amos/Da binding site that contributes to GFP expression in the antennal disc. Mutation of this site reduces GFP expression in the PNCs and the SOPs suggesting that *amos* expression in the PNCs and SOPs is more tightly linked in the case of *amos* than *sc* and *ato*. This result also implies that autoregulation may have an input in regulating PNC as well as SOP enhancers. Future work will continue to investigate these possibilities. In addition to elucidating how *amos* is transcriptionally controlled via *cis*-regulatory elements, confirmation of the mutated site as one to which Amos/Da binds will provide a point of reference to investigate how *amos* specifically regulates its downstream target genes.

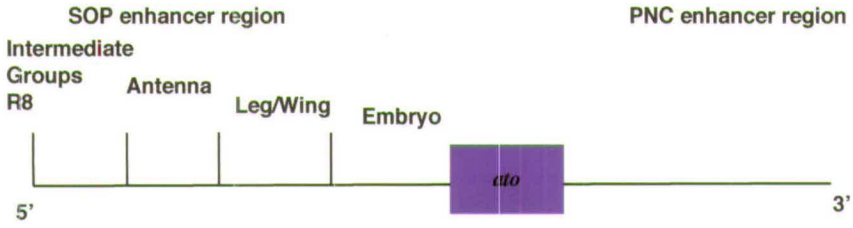


FIG 2.4.1 Schematic of *ato* regulatory regions —1kb

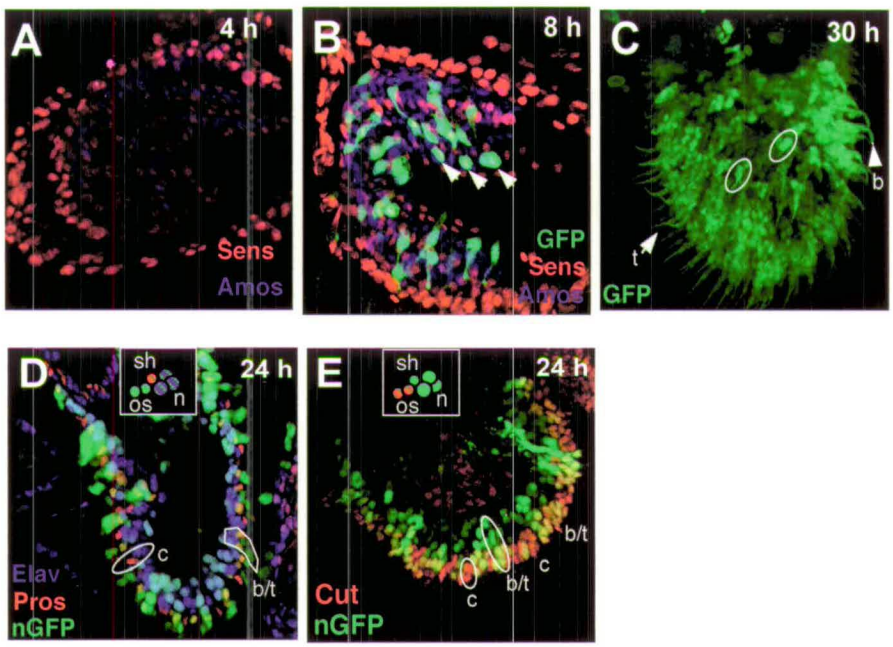


FIG 2.8.1/2 Fate of *amos*- expressing cells during olfactory development. Activity of an *amos* enhancer driving GFP expression in antennal discs - *amosGal4 x UAS GFP* or *UASnlsGFP*. (A) Confocal image of antennal disc 4 h after puparium formation (APF) stained with antibodies to detect Amos protein in blue and Sens protein in red. Sens stains all SOPs. Amos stains both *amos* dependent SOPs and *amos* dependent PNC cells. At 4 h APF, Sens is visible in *amos* (sensilla basiconica and sensilla trichodea) and *ato* (sensilla coeloconica) dependent SOPs, Amos is present in PNC cells, GFP cannot be detected. (B) Confocal image of antennal disc 8 h APF stained with antibodies to detect Amos protein in blue and Sens protein in red. Green is GFP driven by *amos-gal4*. Its is not necessary to use an antibody to visualise the GFP in antennal discs. At 8 h APF, GFP can be detected in amos dependent olfactory precursors, some SOPs are arrowed. (C) Confocal image of antennal disc 30 h APF showing GFP driven by *amos-Gal4*. A large number of sensilla retain GFP. Protein appears to be in sensillar groups (as indicated by rings), and includes the outer support cells, so that sensilla trichodea and basiconica can clearly be discerned (t, b). (D) Confocal image of antennal disc 24 h APF stained with antibodies to detect Elav protein in blue and Pros protein in red. Each sensilla consists of outer support cells (os), neurons (n) and sheath cell (sh). See insert. Elav labels neurons and Pros labels sheath cells. *amosGal4 x UAS GFP* labels rows of cells corresponding to each sensillum basiconicum or trichodeum (*amos*-dependent - some are ringed- b/t) while presumptive coeloconica (*ato*-dependent -some are ringed - c) do not express GFP. Insert shows that all cells of the sensilla basiconicum and sensilla trichodeum i.e. outer support, neurons and sheath cell express GFP. In addition the neurons express *elav* (blue) and the sheath cell expresses *prospero* (red). (E) Confocal image of antennal disc 24 h APF stained with antibody to to detect Cut protein in red. Cut labels outer support cells. Insert shows that all cells of the sensilla basiconicum and sensilla trichodeum i.e. outer support, neurons and sheath cell express GFP. In addition the outer support cells express *cut* (red).

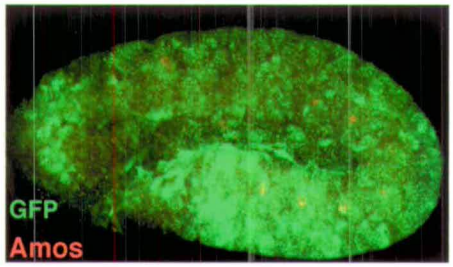


FIG 2.8.3 Expression of *amos-3.6-Gal4* in embryos. Confocal image of stage 11 embryo stained with antibodies to detect Amos protein in red and GFP protein in green. Expression of GFP is widespread and non-specific. The embryo is of genotype *amosGal 4 # 2, UAS GFP.* A similar phenotype was observed with *amosGal 4 # 1, UAS GFP.*

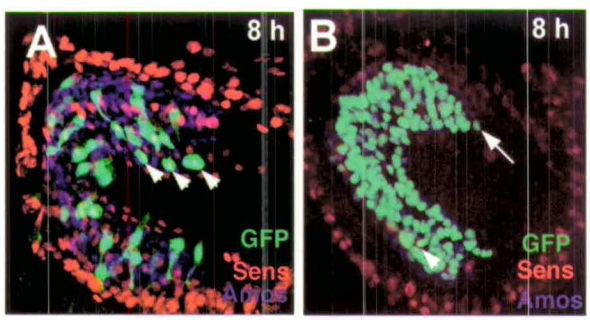


FIG 2.9.1 Comparison between GFP expression driven by *amos*- 3.6-Gal4 and *amos-3.6-GFP* in the antennal disc 8 h after puparium formation. Confocal images of third antennal segment 8h APF stained with antibodies to detect Amos protein in blue and Sens protein in red. Sens staining is observed in all SOPs. GFP expression is driven either by *amos 3.6 Gal4* or *amos 3.6 GFP* and does not require an antibody for visualisation. (A) *amos- 3.6-Gal4 x UAS GFP.* GFP is detected in *amos* dependent SOPs (marked by an arrowhead). *amos* dependent SOPs stain for both Amos and Sens protein. (B) *amos-3.6-GFP* . GFP expression is detected in *amos* dependent SOPs (*GFP. amos, sens*) and PNCs (*GFP, amos*). *amos* dependent SOPs are marked by an arrow head and PNC cells by an arrow. The presence of GFP in PNC cells in B indicates that the enhancer is not solely an SOP one as had been suggested by the expression pattern of *amos-3.6-Gal4, UAS GFP,* rather that the upstream region of *amos* contains enhancer elements for both SOP and PNC cells.

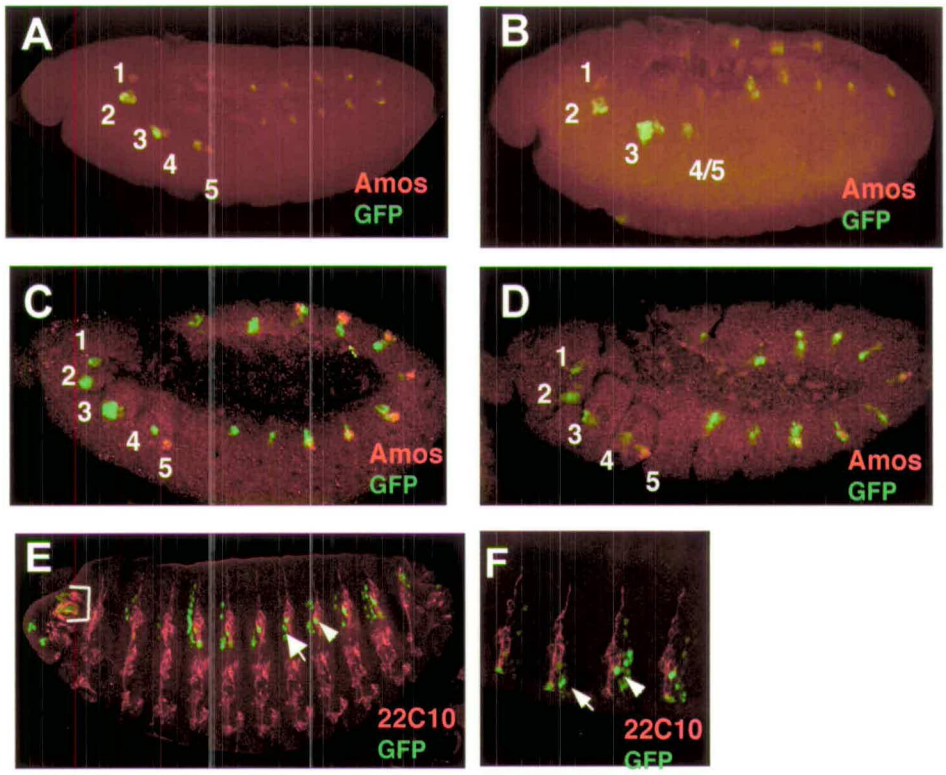


FIG 2.9.2 Expression of *amos-3.6-GFP* in the embryo. Confocal images of embryos of varying ages stained with antibodies to detect GFP protein in green and either Amos protein in red (A-D) or 22C10 protein in red (E,F). (A-B) At stage 10 of embryonic development *amos* is expressed in clusters of cells in the head (clusters 1-5), thoraic and abdominal segments. (A) GFP expression is observed in some cells of clusters 2-4 and also in one or two cells of the clusters in the thoraic/abdominal regions. (B) GFP expression is observed in clusters 2-4 and also a low level of GFP expression is detected in cluster 1. More widespread expression is detected in the clusters of the thoraic/abdominal regions ranging from two to approximately six cells. (C) By stage 11 of embryo development *amos* expression in the head clusters 1-4 has been switched off. *amos* expression remains in cluster 5. GFP driven by *amos -3.6-GFP* is present in clusters 1-4. *amos* expression in the in the trunk has disappeared in the anterior segments and is in the process of being refined to a single cell in the more posterior segments. GFP is present and is mostly located dorsal to the *amos* expressing cells with a small overlap. (D) Late stage 11 embryo, *amos* expression remains in cluster 5. There is also some GFp expression present in this cluster. *amos* expression has either been switched off or rrrefined to a single cell in all thoraic/sbdominal segments. GFP overlaps with the remaining *amos* expression and is observed mainly dorsal to the *amos* expressing cell. (E) Late embryo, 22C10 marks all peripheral neurons, GFP can be detected in the *amos* dependent dmd neuron (marked with an arrowhead)and the dbd neuron and associated glial cell (marked with an arrow) in abdominal segments. GFP is stronger in the more dorsally located dmd neuron. GFP can also be detected in the head region in the antennomaxillary complex (brackets) which are the presumed olfactory organs of the embryo and are thought to be *amos* dependent and in some ectodermal cells. (F) Higher magnification of abdominal segments showing the dmd (arrowhead) and dbd (arrrow) neurons.

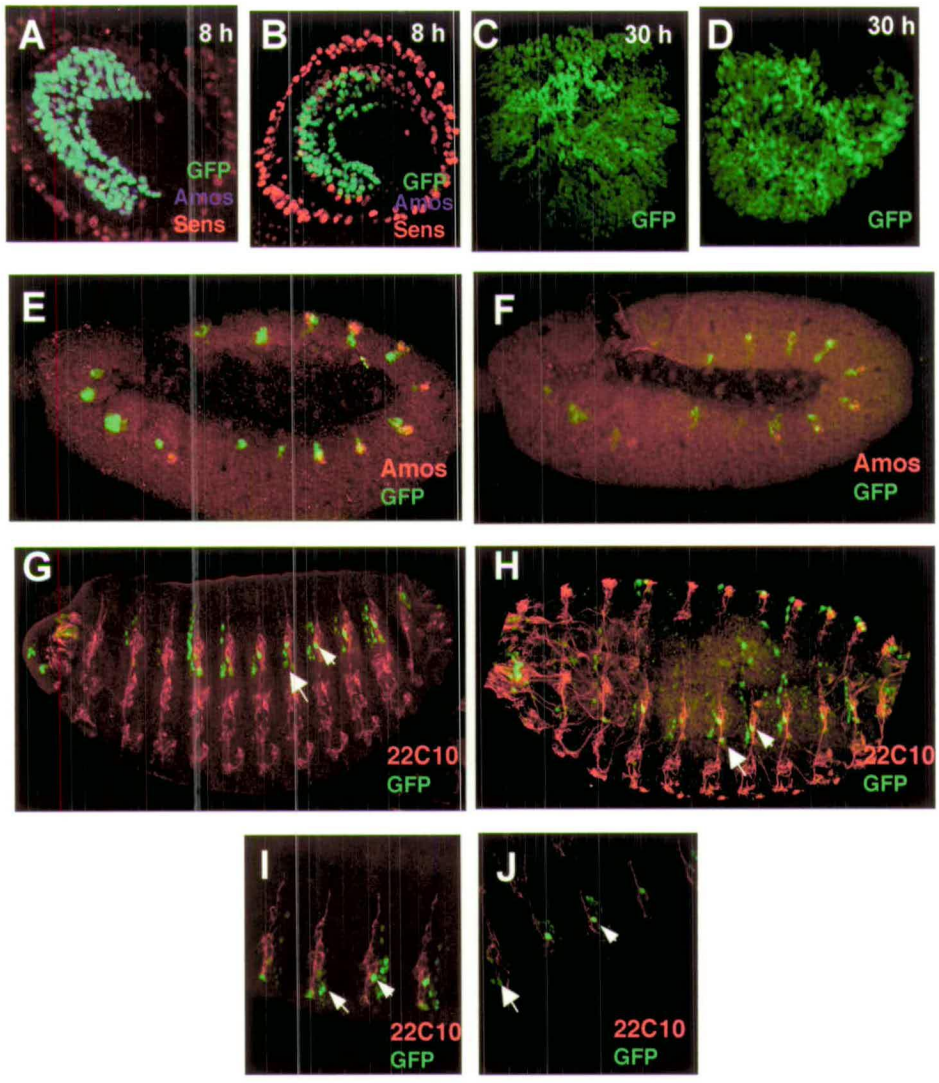


FIG 2.10.1 Comparison between expression of GFP driven by *amos-3.6-GFP 1* and *amos-3.6M GFP .* (A,B) Confocal images of antennal discs 8 h APF stained with antibodies to detect Amos protein in blue and Sens protein in red (C,D) Confocal images of antennal discs 30 h APF. GFP is driven either by *amos 3.6 GFP 1* or *amos 3.6 M GFP 1* and it is not necessary to use an antibody to visualise it in antennal discs. (A) *amos 3.6 GFP 1* (B) *amos 3.6 M GFP 1.* GFP expression in *amos*-dependent SOPs and PNCs of the antennal disc, 8 h APF. (C) *amos 3.6 GFP 1* (D) *amos 3.6 M GFP 1* . Expression of GFP in *amos*- dependent sensilla, the sensilla basiconica and sensilla tricodea 30 h APF. (E,F) Confocal images of stage 11 embryos stained with antibodies to detect Amos protein in red and GFP protein in green. (E) *amos 3.6 GFP 1* (F) *amos 3.6 M GFP 1* . GFP expression overlaps with the *amos* expression in the head and thoraic/abdominal regions. (G-H) Confocal images of later embryos stained with antibodies to detect 22C10 protein in red and GFP in green. 22C10 marks all neurons. GFP expression is present in the *amos* dependent dmd neuron (arrow) and dbd neuron (arrowhead). GFP is also present in some ectodermal cells. (I,J) Confocal images showing a higher magnification of abdominal segments showing the dmd (arrow) and dbd (arrowhead) neurons. There is no discernible difference between GFP expression in flies with the wild type enhancer compared to flies with the mutated one in antennal discs or embryos.

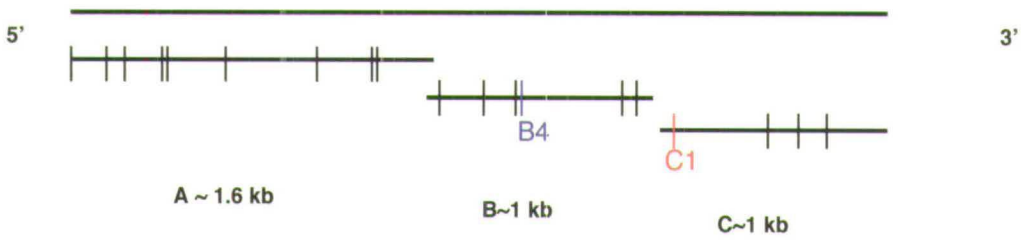


FIG 2.11.1 Subdivision of *amos-3.6* region into three composite fragments, A, B and C. The positions of the E-boxes are marked on the diagram. B4 (blue) and C1 (red) were mutated. See also appendices C and D.

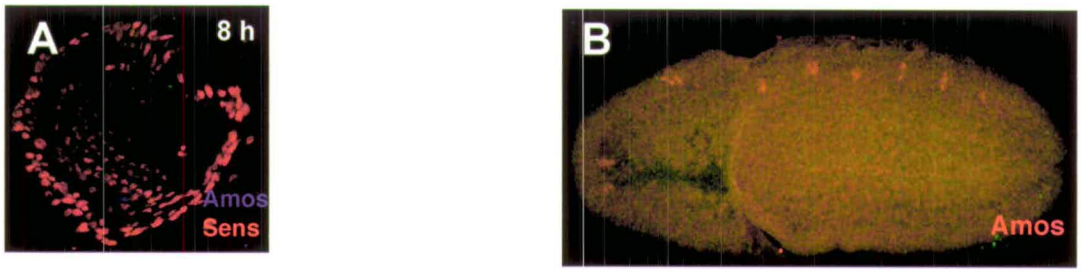


FIG 2.11.2 Expression of *amos-A-GFP* in antennal disc and embryo. (A) Confocal image of antennal disc 8h APF stained with antibodies against Amos protein in blue and Sens protein in red. Unlike *amos 3.6 GFP* (full length fragment) there is no GFP present indicating *amos-A-GFP* does not drive GFP expression in the antennal disc. (B) Confocal image of stage 11 embryo stained with antibodies against Amos protein in red and GFP protein in green. Unlike *amos 3.6 GFP* (full length fragment) there is no GFP present indicating *amos-A-GFP* does not drive GFP expression in the embryo.

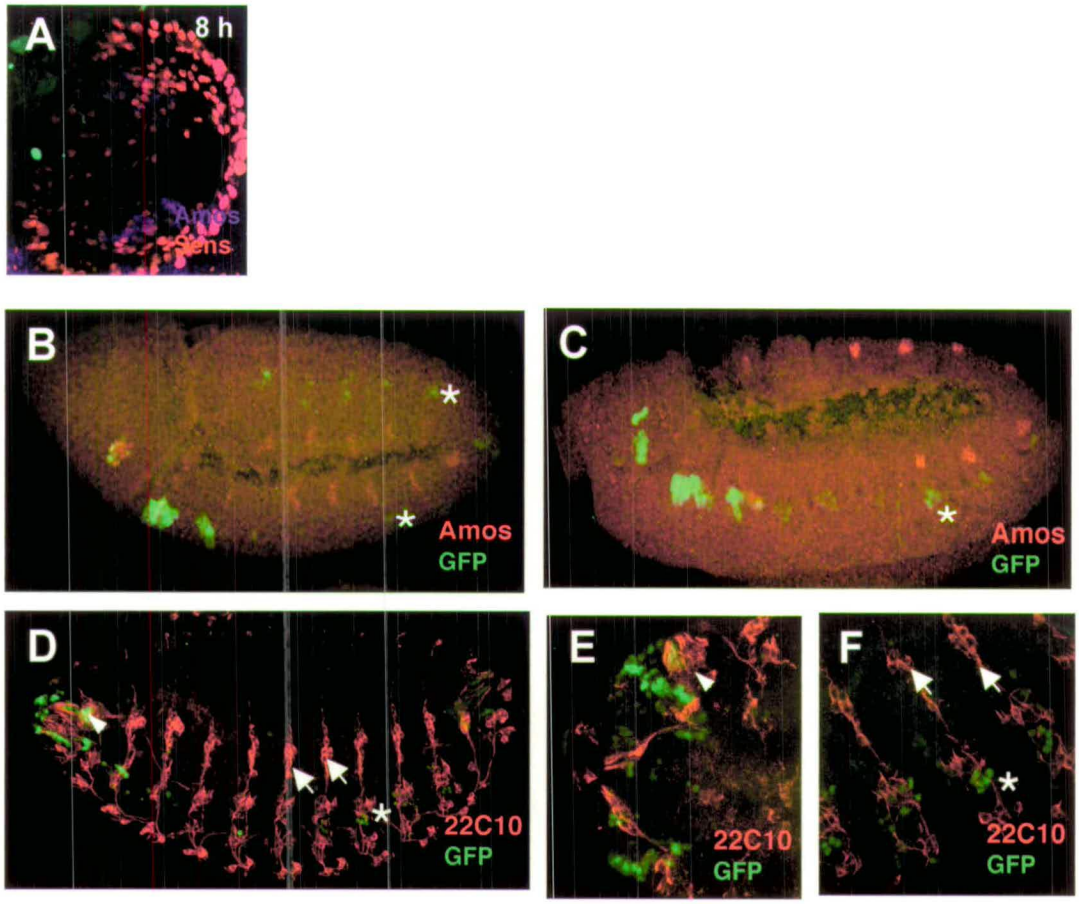


FIG 2.11 .3 Expression of *amos-B-GFP* in antennal disc and embryo. A Confocal image of antennal disc 8 h APF (*amos-B*-GFP) stained with antibodies against Amos protein in blue and Sens protein in red. Unlike *amos-3.6 -GFP* (full length fragment) there is no GFP present indicating *amos-B-GFP* does not drive GFP expression in the antennal disc. (B,C) Confocal images of stage 11 embryos stained with antibodies against Amos protein in red and GFP protein in green. (B) *amos-B-GFP 1.* (C) *amos-B-GFP 2.* GFP is present in the head region in a similar pattern to that of *amos 3.6 GFP* (full length fragment). In addition, ectopic expression is present in the trunk region in clusters larger than that seen with *amos -3.6 - GFP* and is not associated with *amos* staining (*). (D) Confocal image of late embryo (*amos-B-GFP*) stained with antibodies to 22C10 protein in red and GFP protein in green. 22C10 stains all neurons. Expression in the head region is similar to *amos-3.6-GFP* i.e. the antennomaxillary complex (arrowhead) and ectodermal cells. GFP in the trunk region (*) is not associated with the *amos* dependent multiple dendritic neurons, dmd and dbd (arrows). (E) Confocal image of a higher magnification of the head region. The antennomaxillary complex is marked with an arrowhead. (F) Confocal image of a higher magnification of the abdominal segments. GFP staining (*) is not associated with the *amos* dependent md neurons (arrows). *amos-B-GFP* supports expression of GFP in the head regions of the embryo in a similar pattern to that of the full length fragment and therefore seems to contain an enhancer element responsible for expression in the head region. In addition ectopic expression is present in the trunk region in a segmental pattern similar to that of the head region. GFP expression does not appear to be specific to the nervous system.

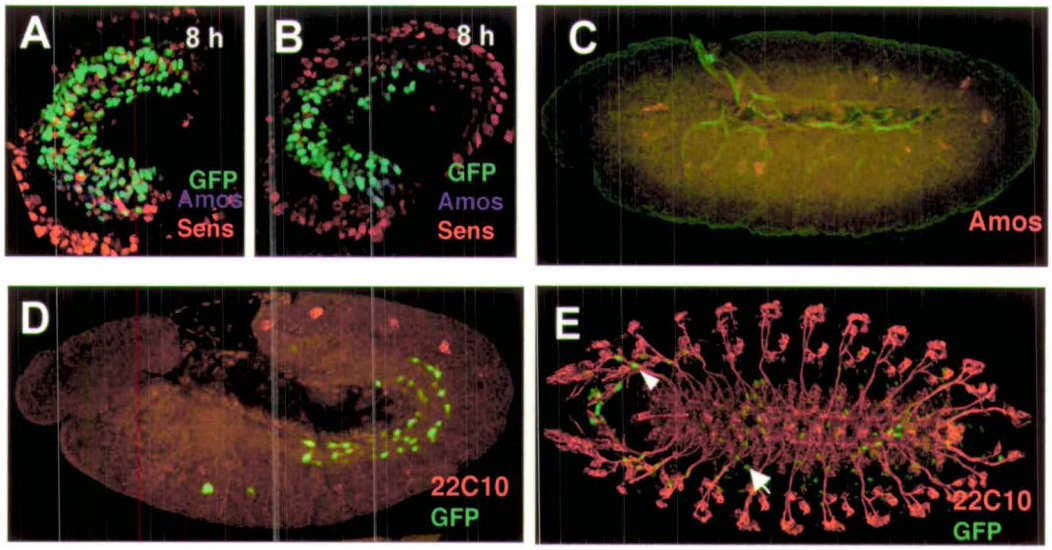


FIG 2.11.4 Expression of *amos-C-GFP* in antennal disc and embryo. (A,B) Confocal images of 8hr antennal discs (*amos-C-GFP*) stained with antibodies to Amos protein in blue and Sens protein in red. Sens stains all SOPs. Green is GFP driven by *amos-C-GFP* and it is not necessary to use an antibody to visualise it in antennal discs. *amos-C-GFP* drives GFP expression in the *amos*-dependent SOPs and PNCs of the antennal disc in a similar pattern to that driven by *amos-3.6-GFP* (full length fragment). (C) Confocal image of stage 10/11 embryo (*amos-C-GFP*) stained with antibodies to Amos protein in red and GFP protein in green. N o GFP is detectable in either the head or trunk regions. (D) Confocal image of late stage 11 embryo (*amos-C-GFP*) stained with antibodies to 22C10 protein in red and GFP protein in green. 22C10 stains all neurons. GFP is present in two rows of internal cells. (E) Confocal image of late embryo (*amos-C-GFP*) stained with antibodies to 22C10 protein in red and GFP protein in green. later embryo, GFP is present in a number of cells, some ectodermal (arrows) and some associated with neurons (arrowheads). The neuron associated cells are not specifically associated with the *amos*-dependent neurons. *amos-C-GFP* appears to contain an enhancer element responsible for *amos* expression in the third antennal segment.

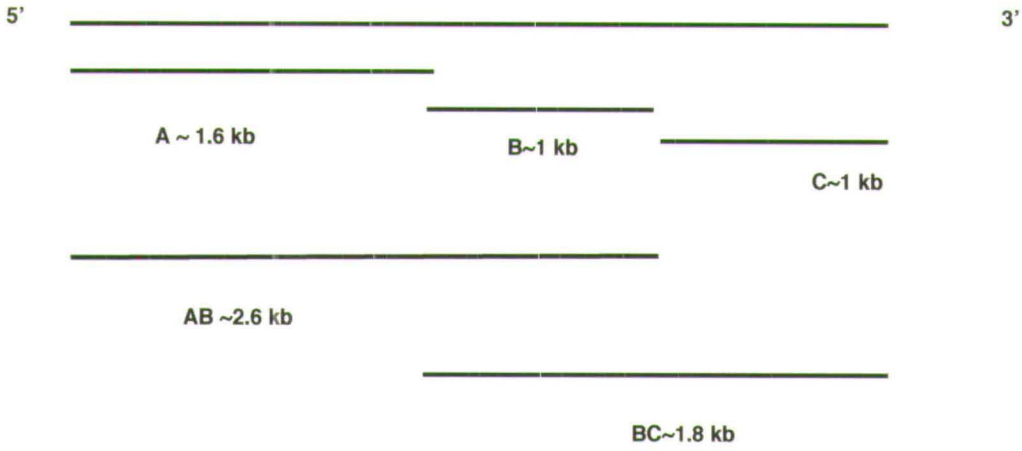


FIG 2.11.5 Schematic of overlapping fragments of *amos* regulatory region

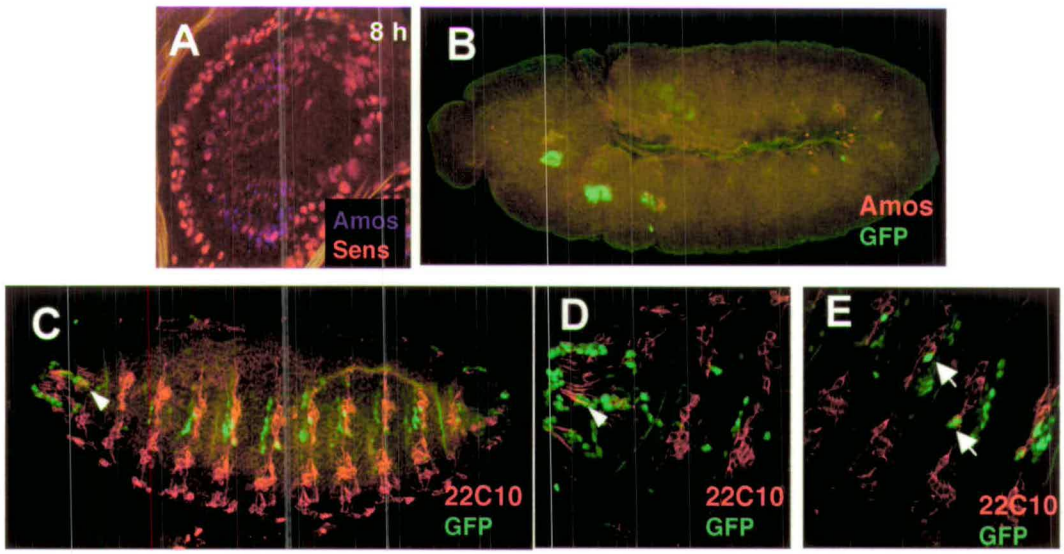


FIG 2.11.6 Expression of *amos-AB-GFP* in antennal disc and embryo. (A) Confocal image of 8h antennal disc (*amos-AB -GFP*) stained with antibodies against Amos protein in blue and Sens protein in red. Sens stains all SOPs. No GFP expression is present in an 8 h antennal disc. (B) Confocal image of stage 10 embryo (*amos-AB -GFP*) stained with antibodies to Amos protein in red and GFP protein in green. GFP expression similar to that observed with *amos-3.6-GFP* (full length fragment) is present in the *amos* expressing cells of the embryo in both the head and trunk regions. (C) Confocal image of later embryo (*amos-AB -GFP*) stained with antibodies against 22C10 protein in red and GFP protein in green. GFP expression is similar to that observed with *amos-3.6-GFP* (full length fragment) i.e. GFP expression in the head region is observed in the antennomaxillary complex (arrowhead) and ectodermal cells. GFP in the trunk region is associated with the *amos* dependent multiple dendritic neurons, dmd and dbd (arrows). (D) Confocal image of a higher magnification of the head region. The antennomaxillary complex is marked with an arrowhead. (E) Confocal image of a higher magnification of the abdominal segments. GFP staining is associated with the *amos* dependent md neurons - dbd and dmd (arrows). *amos-AB-GFP* supports GFP expression in the head and trunk region of the embryo in a similar pattern to that of *amos-3.6-GFP* (full length fragment).

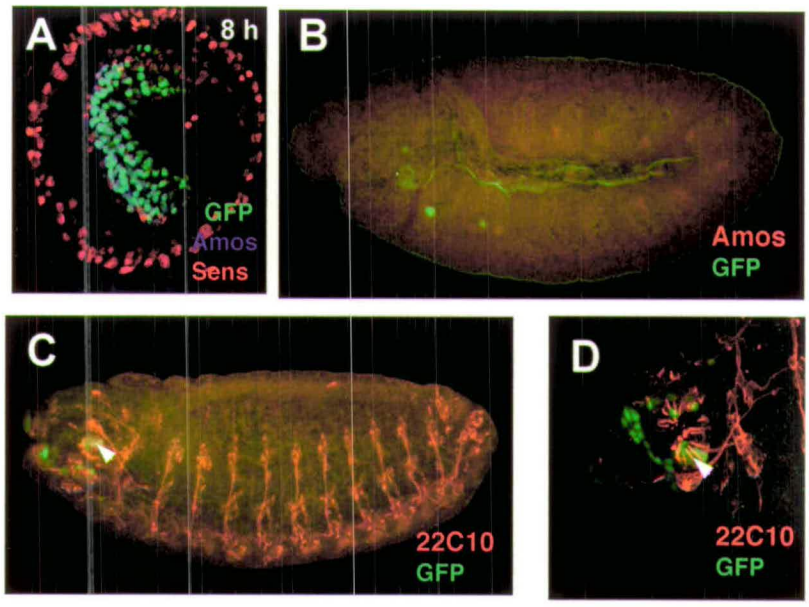


FIG 2.11.7 Expression of *amos-BC-GFP* in antennal disc and embryo. (A) Confocal image of 8h antennal disc (*amos-BC-GFP*) stained with antibodies against Amos protein in blue and Sens protein in red. Sens stains all SOPs. GFP expression is driven by *amos-BC-GFP* and it is not necessary to use an antibody for visualisation. GFP is present in the *amos* dependent SOPs and PNCs of the antennal disc in a similar expression pattern to that observed with *amos-3.6-GFP* (full length fragment). (B) Confocal image of stage 10/11 embryo (*amos-BC-GFP*) stained with antibodies to Amos protein in red and GFP protein in green. GFP expression similar to that observed with *amos-3.6-GFP* (full length fragment) is present in the *amos* expressing cells of the embryo in the head region. Unlike *amos-3.6-GFP* there is no GFP expression in the trunk region. (C) Confocal image of later embryo (*amos-BC-GFP*) stained with antibodies against 22C10 protein in red and GFP protein in green. GFP expression is similar to that observed with *amos-3.6-GFP* (full length fragment) in the head region i.e. GFP expression in the head region is observed in the antennomaxillary complex (arrowhead) and ectodermal cells. There is no GFP expression in the trunk region. (D) Confocal image of a higher magnification of the head region. The antennomaxillary complex is marked with an arrowhead. *amos-BC-GFP* supports GFP expression in the antennal disc and head region of the embryo in a similar pattern to that of *amos-3.6-GFP* (full length fragment).

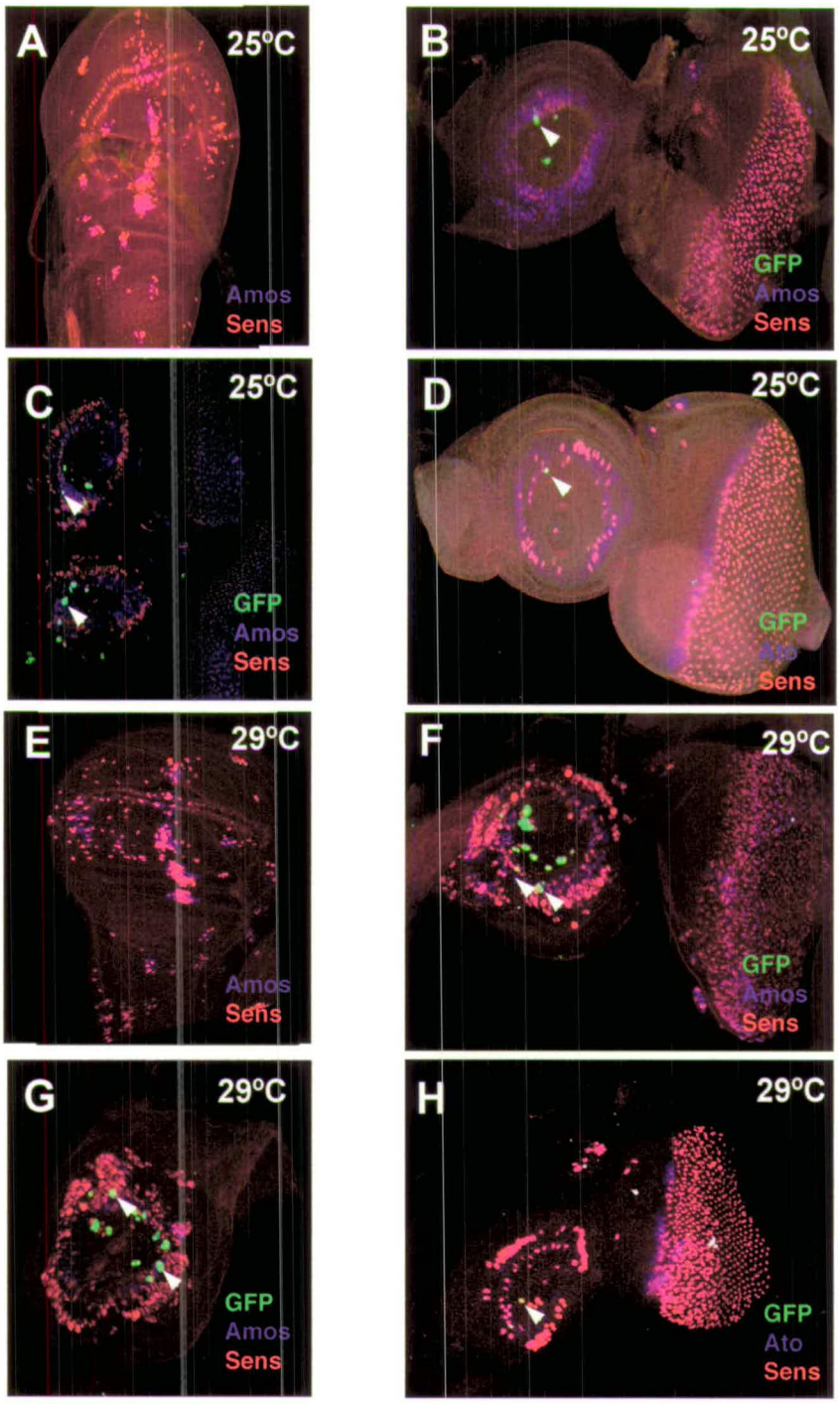


FIG 2.12.1 Misexpression analysis of *amos -3.6 GFP* driven by *109-68Gal4* when crossed to *UAS amos* or *UAS ato* at 25 °C and 29 °C. (A) Confocal image of third instar larval wing disc (*amos-3.6-GFP;109-68Gal4 x UAS amos* at 25°C) stained with an antibody against Amos protein in blue and Sens protein in red. No GFP expression. (B,C) Confocal images of third instar eye antennal discs (*amos-3.6-GFP;109-68Gal4 x UAS amos* at 25°C) stained with an antibody against Amos protein in blue and Sens protein in red. GFP is present in some ectopic SOPs (arrowhead. (D) Confocal image of third instar eye antennal disc (*amos-3.6-GFP;109-68Gal4 x UAS ato* at 25°C) stained with an antibody against Ato protein in blue and Sens protein in red. GFP is present in one some SOP (arrowhead). (E) Confocal image of third instar larval wing disc (*amos-3.6-GFP;109-68Gal4 x UAS amos* at 29°C) stained with an antibody against Amos protein in blue and Sens protein in red. No GFP expression. (F,G) Confocal images of third instar eye antennal discs (*amos-3.6-GFP;109-68Gal4 x UAS amos* at 29°C) stained with an antibody against Amos protein in blue and Sens protein in red. GFP is present in some ectopic SOPs (arrowhead) and ectodermal cells. (H) Confocal image of third instar eye antennal disc (*amos-3.6-GFP;109-68Gal4 x UAS ato* at 25°C) stained with an antibody against Ato protein in blue and Sens protein in red. GFP is present in one some SOP (arrowhead).

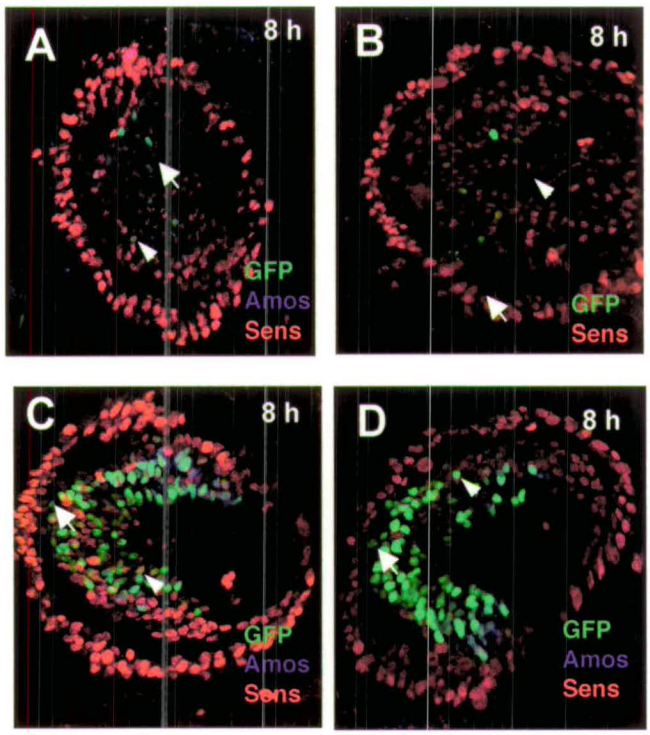


FIG 2.12.2 Comparison between expression of GFP driven by amos-C-GFP and amos-CM-GFP, i.e., mutated and non-mutated amos-C-GFP. (A,B) Confocal images of 8h antennal disc (*amos-CM-GFP*) stained with antibodies against Amos protein in blue and Sens protein in red. Sens stains all SOPs. GFP expression is driven by *amos-CM GFP* and it is not necessary to use an antibody for visualisation. GFP is present in the *amos* dependent SOPs (arrowhead) and PNCs (arrow). (C,D) Confocal image of 8h antennal disc (*amos-C-GFP*) stained with antibodies against Amos protein in blue and Sens protein in red. Sens stains all SOPs. GFP expression is driven by *amos-C-GFP* and it is not necessary to use an antibody for visualisation. GFP is present in the *amos* dependent SOPs (arrowhead) and PNCs (arrow) of the antennal disc. GFP expression is greatly reduced in the mutated fragment when compared to the wild-type. Low-level expression remains in both SOPs and PNCs suggesting the enhancer may play a role in regulating *amos* expression in both SOPs and PNCs of the antennal disc.

3

Microarray analysis to determine the downstream targets of the proneural protein Atonal

3.1 Aims of this chapter

The aim of this section of my PhD was to develop a system that would allow reliable identification of the downstream target genes of the proneural transcription factors. The *Drosophila* peripheral nervous system consists of five principal types of sensory organs. The different subsets of sense organs are clearly very different both in structure and in function yet they all arise in a similar manner, i.e. by groups of undifferentiated cells deciding to become a sense organ. The proneural genes are central to this process. Proneural genes not only confer neural competence for SOP formation but also endow SOPs with neural subtype information.

I am particularly interested in the method whereby neural subtype information is transmitted to specific SOPs. It has been suggested that this specificity is due to selective regulation of different downstream targets (Jarman and Jan, 1995). However, because of the scarcity of known differential target genes this has proved difficult to assess directly. Identification of these genes would allow a picture to

emerge of how the proneural factors function on a global, genomic level. I decided to focus on microarray analysis as the best means of identifying these genes.

3.2 What do Proneural Genes Regulate?

3.2.1 "Generic" target genes

Proneural genes are believed to regulate a common group of "generic" neural target genes for shared neural characteristics and for the lateral inhibition pathway. The components of lateral inhibition, that are suspected to be directly regulated by proneural genes, include *Enhancer of split (E(spl), Bearded (Brd), scabrous(sca),* and *Delta (D).* Generic target genes are also thought to include the zinc finger transcription factor *senseless,* which is implicated in the maintenance of proneural gene expression, and the neural precursor gene *asense.*

Singson *et al.* (1994) investigated the regulation of several genes thought to be "generic" target genes namely *E(spl), Brd* and *sca.* Comparison with previously identified Sc/Da binding sites in other proneural target gene enhancers (Cabrera and Alonso, 1991; Kunisch *et al.*, 1994; Van Doren *et al.*, 1992), allowed a common E-box motif to be pinpointed - G CAGSTG G/T.

Singson *et al.* (1994) showed *E(spl)m4, E(spl)m7* and *E(spl)m8* are expressed in the wing and eye-antennal discs. The downregulation of *E(spl)* appears to permit SOP development (Culí and Modolell, 1998; Jennings *et al.*, 1995; Oellers *et al.*, 1994). The authors also studied other genes implicated in lateral inhibition; *sca* (Mlodzik *et al.*, 1990) and *Brd* (Lai *et al.*, 2000). *sca* and *Brd* are expressed in a

proneural cluster pattern. In flies of the genotype sc^{10-1} (inactivates *ac* and *sc)*, expression of these genes is abolished in the wing disc but is unchanged in the eye-antennal disc suggesting that expression in the wing disc is under the control of the *AS-C* genes and that expression in other regions is independent of *AS-C*. E-boxes were identified in the upstream regions and DNA binding studies showed that heterodimers of Da/Sc and Da/Ac bind strongly to these sites *in vitro*. Mutation of the E-boxes prohibits binding, which suggests that members of the *AS-C* regulate these lateral inhibition components through these E-boxes. Fragments of the *E(spl)* and *Brd* promoters that include the E-boxes are sufficient to drive *ac-sc* dependent expression of a reporter gene in the wing disc proneural clusters. Mutation of the E-box sites abolishes activation of the promoters in the proneural clusters of the wing disc but not in the eye-antennal disc suggesting expression in the eye-antennal disc is independent of *AS-C* (Bailey and Posakony, 1995; Singson *et al.*, 1994).

Other "generic" target genes include *asense (ase)* and *senseless (sens). ase* is expressed in all neural precursors of the CNS and PNS in embryos and imaginal discs and occurs after expression of the proneural genes has been switched off (Brand *et al.*, 1993). *ase* expression in a *sc* mutant background leads to abolition of *ase* in the *sc*-dependent SOPs of the embryo but expression remains, among others, in the chordotonal organs and the dbd neuron, suggesting *ase* is regulated by the all proneural genes. (Jarman *et al.*, 1993a) showed that the regulatory element of *ase* possesses AS-C/Da binding sites (E1-CAGCTG and E4 -CAGGTG), and that deletion of these sites reduces expression of a fusion gene. As expression is reduced but not abolished, this suggests other factors may also regulate *ase*.

sens is a zinc finger transcription factor that is activated by proneural gene expression. Sens is then in turn required to further activate and maintain proneural gene expression (Nolo *et al.*, 2000). The upstream region of *sens* contains a number of E-boxes, including one of sequence G CAGGTG, the preferred binding site for Sc/Da heterodimers (Singson *et al.*, 1994). That proneural proteins are capable of binding to this site is suggested by the ability of the upstream region to drive *lacZ* expression in a pattern reminiscent of endogenous *sens* expression. A very recent paper has shown that that proneural proteins directly regulate *sens* expression (Jafar-Nejad *et al.*, 2003). Mutation of the CAGGTG E-box results in a reduction of reporter gene expression in the embryo and pupae. The fragment tested contained only one E-box and mutation of this site caused expression to be almost completely abolished in all the PNS cells of the embryo. This suggests the different proneural proteins may bind the same site *in vivo*.

These genes in most cases have been identified as direct downstream targets of *sc* and *ac*. The assumption up to recently has been that the proneural gene *ato* and, possibly, *amos*, also regulate at least some of these genes. This is because these "generic" target genes are involved in processes common to all proneural genes. Moreover, this has been thought to occur through shared E-boxes with tissue-specific protein cofactors providing the specificity.

3.2.2 Differential Target Genes

The suggestion that proneural specificity is due to selective regulation by different downstream target genes (Jarman and Jan, 1995) has proved difficult to

assess directly due to the absence of known differentially regulated target genes. There are some candidates for *ato,* namely *rhomboid, NKD* and *cato. rhomboid (rho),* which is expressed in all *ato*-dependent neural precursors is activated by *ato* leading to EGFR activation. Ip *et al.*, (1992) analysed *rho* regulation in relation to EGFR signalling in dorsoventral patterning and work in the Jarman Lab has extended this to *rho* regulation in neurogenesis.

Another candidate Ato target is the tachykinin receptor related gene, *TAKR86C* (formerly *NKD*). Rosay *et al.* (1995) showed, via reporter gene constructs, that the proximal region of the promoter is responsible for expression in the dorsal SOP of the posterior group or P cell. These are the precursors of the lateral chordotonal organ in the abdominal segments (Jarman *et al.*, 1993b). The authors showed that an Ato/Da heterodimer binds to the TAKR86C promoter via an E-box of sequence (T CAGGTGT). Powell *et al.*, (submitted) constructed artificial multimers of the E-box site. The resulting construct supported expression in a subset of *ato*-dependent SOPs.

A third potential target of *ato* is *cato,* which encodes a protein with a bHLH domain that is closely related to that of Ato (Goulding *et al.*, 2000a). The early phase of *cato* expression is highly specific for *ato*-dependent chordotonal precursors and is activated immediately *ato* expression is refined to these cells (Goulding *et al.*, 2000a). The presence of E-boxes in the promoter region of *cato* suggests *ato* may regulate *cato* in a similar manner to that described above but this has not been explored.

There are no known target genes of *amos*. It has been assumed that *amos* may regulate some of the generic target genes like *ase* and *E(spl)C* but this is based purely on speculation. In addition, it is thought *amos* must regulate differential downstream target genes to confer subtype specificity. The subject of *amos* specific downstream target genes remains a black box. In an attempt to open up an avenue for exploration I have studied the *cis*-regulatory elements of *amos*. Identification of the E-box responsible for autoregulation would provide a point of reference to compare other putative downstream target genes. (See Chapter 2).

All the "generic" target genes identified thus far have been isolated on the basis of regulation by AS-C. These genes appear to regulate common aspects of neurogenesis such as lateral inhibition and, while it is on this basis that they are believed to be also targets of other proneural genes, in fact the only case where this has been shown to be correct is *Brd* which is regulated by Ato as well as by AS-C. There is no indication that any of the AS-C downstream target genes identified thus far, are regulated exclusively by AS-C. This is in contrast to *ato,* where *TAKR86C, rho* and *cato* appear to be regulated exclusively by Ato. Therefore it is possible that *sc* may regulate so called "generic" target genes only and it is the additional selective regulation of *ato* and *amos* that gives rise to proneural specificity. This is in agreement with the results of misexpression assays. Misexpression experiments have shown that *amos* and *ato* can impose olfactory or chordotonal fate, respectively, on external sense organs (Goulding *et al.*, 2000b; Jarman and Ahmed, 1998). However the reverse is not true i.e. the genes of the *AS-C* are unable to impose an external sense organ fate on *amos* or *ato* dependent precursors. This suggests *amos* and *ato* regulate more specific aspects of neural development than *AS-C*. Or to put

it another way, formation of external sense organs is the "default" state, and formation of olfactory and chordotonal organs require another level of complexity involving differential downstream target genes.

The proneural proteins are nodal control points in neurogenesis and increased knowledge of the downstream targets of these genes will shed light on how neural subtype information is transmitted to each SOP. If, as is suspected, *amos* and *ato* regulate more specific aspects of neurogenesis than *sc*, this should be reflected in their respective downstream targets. I chose to study this at the whole genome level and have piloted the use of microarray technology to determine the downstream target genes of *ato* and carried out preliminary analysis to test the validity of this approach. Known target genes, both "generic" and differential, were assessed to determine if there was a change in expression between sample and control. In addition a number of new putative downstream targets of *ato* were identified. Before deciding on microarray analysis as the best way to identify downstream targets, I considered a number of different options.

3.3 Possible methods of identifying differentially regulated target genes

A number of methods of differential screening were considered. One possible approach would be to construct precursor specific cDNA libraries. The method used by Dulac and Axel (1995) in their attempt to isolate pheromone receptor genes was to construct a cDNA library from single neurons for differential screening. This method gave an unbiased representation of a given RNA. This was also one of the

methods used by Bryant *et al.* in their 1999 paper to study differentially expressed genes in *Drosophila* follicle cells.

Differential display (Liang and Pardee, 1992) would provide an alternative method for differential screening. This would involve using a set of oligonucleotide primers, one of which is anchored to the polyadenyl tail of a subset of the mRNA, the other being short and arbitrary in sequence to facilitate annealing different positions relative to the first primer. mRNA subpopulations defined by the primer pairs are amplified after reverse transcription and resolved on a DNA sequencing gel. The advantages of this method when compared to subtractive and differential hybridisations are that it is considered more sensitive and is much quicker as it does not require cDNA library construction.

A third possibility would be to utilise a technique known as representational difference analysis (RDA) (Hubank and Schatz, 1994). This is a process of subtractive hybridisation coupled to amplifications that was originally used for identifying the difference between genomic DNA from two genomes. The advantage of RDA when compared to differential display is that the latter amplifies fragments from all represented mRNA species, whereas RDA eliminates those fragments present in both populations leaving only the differences. This technique has been found to be faster, extremely sensitive, reproducible and predominantly lacking false positives. It has been used to identify Pax 7 as a gene specifically expressed in mouse satellite derived myoblasts (Seale *et al.*, 2000).

3.4 Microarray analysis

A fourth approach considered was microarray analysis. Given that the *Drosophila* genome has been sequenced, this option was considered the most attractive. Microarray analysis has been used with some success to identify differentially expressed genes in *Drosophila* follicle cells (Bryant *et al.*, 1999), to study *Drosophila* development during metamorphosis (White *et al.*, 1999), to investigate the patterns of gene expression during *Drosophila* mesoderm development (Furlong *et al.*, 2001) and to study dorsal-ventral patterning in the embryo (Stathopoulos *et al.*, 2002). Two possible avenues were explored, one of which was the service offered by the BBSRC functional genomics facility at the University of Cambridge. The second avenue involved using commercially produced Affymetrix gene chips readable in either the Scottish Centre for Genomic Technology at the University of Edinburgh or the Sir Henry Wellcome Genomics microarray facility, which is located at the University of Glasgow. The chips used by the BBSRC microarray facility consisted of the Drosophila Gene Collection (DGC) EST set generated by the Berkley Drosophila Genome Project. This unique set of 5,600 EST's consists of approximately 40% of the genes in the *Drosophila* genome. Given the availability of Affymetrix chips containing probe sets for the more than 13,500 genes (the entire *Drosophila melanogaster* genome), the commercially available Affymetrix technology was by far the better option.

3.4.1 Principle behind microarrays

The principle upon which microarrays are based is hybridisation. Oligonucleotide probes are immobilised on the chip surface, micrometers apart and

located at very specific points. Messenger RNA is extracted from the sample of interest, fluorescently labelled and hybridised to the chip. The fluorochrome is then used to detect hybridisation between the target mRNA and a specific probe on the chip. The chip is scanned in a laser scanner and the image analysed by computer. When different probes matching all the mRNAs in the sample are used, the expression profile of that sample at a particular time is reflected. This is an extremely powerful means of determining which genes are switched on at a given moment in time. Loss - of- function and gain- of-function analysis of a particular gene compared both with each other and with wild type, can be used to determine which genes are under the control of a particular transcription factor. Microarray technology allows the simultaneous monitoring of thousands of genes.

The Affymetrix GeneChip system, uses prefabricated oligonucleotide chips and the spotting of each probe is very uniform. Of the 13,500 genes present on the *Drosophila* chip, over 8,000 of the genes represented have at least one EST/cDNA match. The remaining 5,500 sequences had a gene predicted algorithmically. Affymetrix chips are made by synthesising oligonucleotides on the surface of a chip. Sets of oligos for each gene are present on the chip. In the case of the *Drosophila* chip, 28 oligos per transcript are present. Each oligo is 25 nucleotides long and has been designed to a region of the transcript that is dissimilar to other genes. Each oligo is of a unique sequence. Fourteen of the oligos are exactly complementary to the mRNA of the gene and are termed perfect match (PM). The remaining fourteen have one nucleotide changed to its complementary nucleotide and are called mismatch oligos (MM). The MM probes are present in order to detect what percentage of hybridisation is due to cross hybridisation and as such they act as an

internal control. The Affymetrix chips use only one fluorochrome per chip and so two chips are required to compare sample and control.

A basic outline of the procedure used to obtain labelled mRNA is as follows.

1. Isolation of RNA using either Trizol or RNeasy kits.
2. First strand cDNA synthesis using reverse transcriptase and an oligo dT primer.
3. Second strand cDNA synthesis and clean up of double-stranded cDNA
4. *In vitro* transcription (IVT) to synthesise cRNA using T7 RNA polymerase in the presence of biotin-UTP and biotin-CTP. Clean up of biotin-labelled cRNA.
5. Fragmentation of cRNA to produce cRNA fragments of length 35 to 200 nucleotides.
6. Hybridisation to chip, followed by washes to remove non-hybridised material. The hybridisation cocktail contains spikes from other organisms, namely bioB, bioC and bioD from *E. coli* and cre from P1 bacteriophage. These spikes are used as hybridisation controls.
7. Stain hybridised biotin-labelled cRNA with Streptavidin-Phycoerythrin, wash, add goat IgG and biotinylated antibody to amplify the signal on the chip.
8. Scan chip in scanner.

3.5 Applications of Microarray Technology

Microarray technology has been used extensively in expression analysis involving many organisms. Affymetrix chips are available for many different

species including human, rat, mouse and *C. elegans* as well as *Drosophila.* This technology can be used to determine the presence of a particular RNA at a particular moment in time and to study the changes in RNA expression patterns over time. It can be used to study human disease to determine how a particular disease affects the expression of all genes in the cell. In addition, microarray analysis can be used in genotyping i.e. detecting mutations in specific genes.

3.5.1 Microarrays in *Drosophila.*

In recent years there has been a huge increase in studies involving microarray technology in all organisms including *Drosophila.* These experiments have involved many different developmental processes in a number of different tissues. However, they can be divided two broad categories:

(i) those that focus on a developmental process to determine the changes in gene expression over time (Arbeitman *et al.*, 2002; White *et al.*, 1999)

(ii) those that attempt to discover downstream targets of a particular transcription factor involved in processes such as mesoderm development, (Furlong *et al.*, 2001), gliogenesis (Egger *et al.*, 2002) or dorsal-ventral patterning in the embryo (Stathopoulos *et al.*, 2002).

My experiments fit into category (ii), i.e. I wished to determine the downstream targets of the bHLH transcription factors *sc, ato* and *amos* with a view to elucidating their mechanism of specificity. I then hoped to study the putative downstream targets; firstly, to determine what they are and secondly, their mode of action.

One of the first highly significant microarray papers focusing on *Drosophila* was concerned with development during metamorphosis (White *et al.*, 1999). The authors constructed custom arrays containing several thousand *Drosophila* genes using cDNA EST clones from the Berkeley *Drosophila* Genome Project (BDGP). They then extracted RNA from six different developmental stages (ranging from 18 h before puparium formation to 12 h after) to obtain an insight into the changes in gene expression that occur during metamorphosis. This resulted in identification of a number of genes that may be involved in the process of metamorphosis. Arbeitman *et al.* (2002) used a custom array to analyse the RNA expression levels of 4,028 genes in wild type flies from fertilization to ageing adults. This analysis provided valuable information about the transcriptional programs that underlie the life cycle and compared development of males and females. De Gregorio *et al.* (2001) used Affymetrix technology to identify new genes involved in the immune response in *Drosophila*. The authors monitored the change in gene expression in adult flies in response to microbial infection and by this means identified 400 genes that may play a role in the immune response.

In addition to extracting RNA from the varying different life-cycle stages, i.e. embryos, larvae, pupae or adult flies, microarray analysis has also been carried out on imaginal discs rather than the whole larvae. Custom array technology has been used to investigate the expression profile of *Drosophila* imaginal discs by comparing discs with larval tissues and, also, different types of discs to each other. This resulted in the identification of genes that are expressed in specific imaginal discs (Klebes *et al.*, 2002). Butler *et al.* (2003) dissected wings discs into two complementary regions: the presumptive wing/hinge and the presumptive body wall

One of the first highly significant microarray papers focusing on *Drosophila* was concerned with development during metamorphosis (White *et al.*, 1999). The authors constructed custom arrays containing several thousand *Drosophila* genes using cDNA EST clones from the Berkeley *Drosophila* Genome Project (BDGP). They then extracted RNA from six different developmental stages (ranging from 18 h before puparium formation to 12 h after) to obtain an insight into the changes in gene expression that occur during metamorphosis. This resulted in identification of a number of genes that may be involved in the process of metamorphosis. Arbeitman *et al.* (2002) used a custom array to analyse the RNA expression levels of 4,028 genes in wild type flies from fertilization to ageing adults. This analysis provided valuable information about the transcriptional programs that underlie the life cycle and compared development of males and females. De Gregorio *et al.* (2001) used Affymetrix technology to identify new genes involved in the immune response in *Drosophila*. The authors monitored the change in gene expression in adult flies in response to microbial infection and by this means identified 400 genes that may play a role in the immune response.

In addition to extracting RNA from the varying different life-cycle stages, i.e. embryos, larvae, pupae or adult flies, microarray analysis has also been carried out on imaginal discs rather than the whole larvae. Custom array technology has been used to investigate the expression profile of *Drosophila* imaginal discs by comparing discs with larval tissues and, also, different types of discs to each other. This resulted in the identification of genes that are expressed in specific imaginal discs (Klebes *et al.*, 2002). Butler *et al.* (2003) dissected wings discs into two complementary regions: the presumptive wing/hinge and the presumptive body wall

region. The authors used Affymetrix technology to identify a set of 94 transcripts enriched in the body wall regions and 56 in the wing/hinge region. The *in situ* hybridisation patterns of 50 previously uncharacterised genes correlated with the transcript enrichment identified by the array analysis.

The second broad application of microarray analysis is to determine the transcriptome changes that occur after genetic manipulation of a specific gene. Leemans *et al.* (2001) used custom array technology to determine the downstream target genes of the transcription factor *labial* in the *Drosophila* embryo. *labial* is a homeotic gene and the most proximal gene of the Antennapedia complex. Using a heat-shock promoter, the authors overexpressed *labial* in embryos and found significant expression level changes for 96 genes. Furlong *et al.*, (2001) studied the patterns of gene expression during mesoderm development. Since the bHLH transcription factor Twist initiates mesoderm development, transcriptional profiles of *twist* loss-of-function embryos were compared with embryos in which *twist* was ubiquitously expressed with wild type embryos. This analysis resulted in the identification of hundreds of genes that may play a role in mesoderm development. RNAi was carried out to assess the role in mesoderm development of the zinc finger transcription factor *gleeful*. Injection of *gleeful* dsRNA caused severe loss and disorganisation of somatic muscle cells but did not affect heart and visceral muscle.

Another important transcription factor in *Drosophila* is *glial cells missing (gcm)*. *gcm* is the switch between neuronal and glial fate and is thought to activate glial cell fate through its action on downstream target genes. Egger *et al.*, (2002) analysed differential gene expression in wild-type embryos compared with embryos

in which *gcm* is misexpressed throughout the neuroectoderm. Tissue-specific expression was achieved by using a *scabrous-GAL4* line. RNA was extracted from embryos at two different stages, firstly at stage 11 to attempt to identify direct downstream targets of *gcm* and also at stage 15/16 to identify additional indirect targets. At stage 11, over 400 genes were differentially expressed, and this figure rose to 1,200 by stage 15/16. At both stages, glial specific genes were upregulated and neuron specific genes downregulated. RNA *in situs* were carried out for a subset of the differentially regulated genes to confirm the array data.

Another developmental process, which has been analysed from the point of view of downstream targets of a specific gene, is dorsal-ventral patterning in the embryo. Array analysis has been carried out on the Rel transcription factor Dorsal (Stathopoulos *et al.*, 2002). The Dorsal gradient initiates the differentiation of the mesoderm, neurogenic ectoderm, and dorsal ectoderm across the dorsal-ventral axis of the embryo. Stathopoulos *et al.*, (2002) extracted RNA from early mutant embryos encompassing those that contain either no Dorsal protein, uniformly low levels of Dorsal or uniformly high levels of Dorsal throughout the embryo and then hybridised the samples to Affymetrix chips. This identified 40 new Dorsal target genes, including cell signalling proteins and transcription factors. *In situ* expression patterns were used to validate the array data. The significant advance in this study is that, tissue-specific enhancers were identified for new Dorsal target genes and bioinformatics were employed to identify conserved *cis*-regulatory elements that respond to similar thresholds of the Dorsal gradient. In addition to identifying known motifs three novel shared motifs were also recognised. This is exactly the type of bioinformatic analysis I would like to carry out to identify proneural binding

sites and also other shared motifs to determine how the target genes are regulated. Given the high degree of success experienced by others when using micorarray analysis to identify target genes, I decided this approach was a suitable one to apply to the proneural genes.

3.6 Experimental approach

The basic question I wished to answer was how do proneural factors function on a global, genomic level. Therefore, I focused on microarray analysis as a possible method for reliable identification of putative downstream target genes. I decided to pilot the approach of microarray analysis by concentrating on the proneural gene *ato*. Initially, I intended to compare wild type with mutant embryos but later I expanded my strategy to compare wildtype embryos with mutant embryos, and with embryos in which *ato* is misexpressed throughout the neuroectoderm, i.e. wild type with both loss-of-function and gain-of-function. The ideal outcome of this experiment would be the downregulation of a particular gene in the mutant when compared to wild type, and the upregulation of that same gene in *scabrousGal4 x UAS-ato* when compared to wild type. Of course the reverse is also possible i.e. upregulation in the mutant and downregulation in *scabrousGal4 x UAS-ato*.

I chose *ato* for the following reasons.

(i) The mutant (ato^{1}) is viable and so embryos can be easily collected. RNA and protein are still produced in proneural clusters but there is no subsequent refinement to single sense organ precursor cells. This is because there is a point mutation in the DNA binding domain rendering

ato^{1} an amorphic mutation (Jarman *et al.*, 1995). *sc* mutants are homozygous lethal and so it would be necessary to hand-select embryos using a green fluorescent protein (GFP) balancer. Balancer chromosomes are used to select homozygous mutant progeny by the absence of markers that are present on the balancer chromosomes. Casso *et al.* (1999) constructed GFP expressing balancer chromosomes that can be recognised in living embryos. Hand selected embryos could then be stored in Trizol at –80 °C until sufficiently high numbers are obtained to make RNA.

(ii) Expression of *ato* during embryogenesis should be sufficiently widespread to allow detection of significant changes in gene expression in mutant compared to wild type. *amos* mutants are viable (zur Lage *et al.*, 2003) but *amos* is expressed very transiently in a small number of cells during embryogenesis (Goulding *et al.*, 2000b) and detection of significant gene expression changes may be therefore beyond the capabilities of microarray analysis when using whole embryos. Therefore *ato* was by far the best candidate to use in microarray analysis.

3.6.1 Optimisation of embryo staging

I determined the optimum time of *ato* expression in sense organ precursor cells in the embryos by antibody staining and RT-PCR. Embryonic expression of *ato* is initially in the proneural clusters in the abdominal and thoracic segments before being resolved to four or five cells in each segment. These are the precursors of the embryonic chordotonal organs (Jarman *et al.*, 1993b). The point at which expression

is refined to SOPs is the time point that I wish to use in my comparison of wild type and mutant embryos. This is because my aim is to determine the direct downstream targets of *ato* that allow specification of *ato*-dependent SOPs.

Total RNA can be extracted by a variety of means. I initially extracted RNA using Trizol. Trizol is mono-phasic solution of phenol and guanidine isothiocyanate which maintains the integrity of the RNA during sample lysis. I carried out a two-step RT-PCR, using oligo-dT primers in the RT step and gene-specific primers in the PCR step. Before carrying out the time course experiment, I extracted RNA from an overnight collection of wild type (*Oregan R*), embryos and performed RT-PCR using primers to *echinoid* (*ed* - widely expressed cell adhesion molecule- (Bai *et al.*, 2001), *sens* and *ato*. In each case I obtained a band of expected size and there was no genomic DNA contamination.

Therefore, I next extracted RNA from wild type (*OrR*) and *ato* mutant embryos with the aim of determining the optimum time of *ato* expression. RNA is still produced in the *ato* mutant but expression is not resolved to a single cell. Although this approach is not a quantitative one - in order to determine accurately the amount of RNA present I would need to carry out quantitative RT-PCR - it helped in determining the most suitable time point. By about six hours expression should be resolved to single cells. I concentrated on extracting RNA from time points between 6.0 and 7.5 hours after egg laying. However in all cases, I had difficulty in obtaining a complete picture of *ato* expression as can be seen in fig 3.6.1. But in general *ato* expression can be observed at most of the time points for both wild type and mutant embryos.

FIG 3.6.1

In a parallel approach, I carried out stainings using an antibody against Ato of whole mount staged embryos. Between 6.5 and 7.5 hours, expression was refined to four single cells in each abdominal segment whereas, at the slightly earlier time point, this was not true in all embryos looked at. Therefore I decided embryos aged between 6.5 and 7.5 hours were the most suitable for use in microarray experiments.

FIG 3.6.2

3.6.2 RNA Extraction for array experiments

Having established the embryo staging to use, I now determined the quantity of embryos necessary to obtain sufficient RNA for microarray analysis. The recommended amount of total RNA for microarray experiments was 20 μg at this time. Using Trizol, I determined that an average of 0.069g of embryos yielded 27.38 μg of RNA. This is more than sufficient for a microarray experiment. The quality of total RNA prepared was assessed by the ratio of absorbance at 260 to 280 nm (A_{260} / A_{280}). A ratio of between 1.8 and 2.0 was always obtained, indicating that the RNA was of high quality.

In addition to Trizol, I assessed a kit available from Sigma called Genelute total RNA kit. This kit is a column-based method and allows rapid extraction of RNA in about 30 minutes. In contrast, RNA extraction using Trizol tends to take over an hour. From an average of 0.060g of embryos I obtained an average of 38.9 μg of RNA and the ratio of extracted RNA was always between 1.8 and 2.

Due to the shorter time needed to extract RNA when using the kit compared to Trizol, and because of the importance attached to ensuring all traces of Trizol have been removed, I decided that the kit was the most suitable means of extracting RNA for use in microarray experiments.

3.6.3 Microarray experiments carried out at the Scottish Centre for Genomic Technology

My initial experiments were carried out at the Scottish Centre for Genomic Technology (SCGT). As will be seen, these were ultimately unsuccessful and a brief description follows below.

I extracted RNA from wild type and mutant embryos between 6.5 and 7.5 hours after egg laying and transported the RNA to the SCGT on ice. I was able to obtain sufficient embryos to extract 20 μg of RNA from single collections. This meant I did not have to store embryos at –80 °C. The samples were checked on an Agilent bioanalyzer for RNA quality. This is termed "lab on a chip" and is a capillary-based system whereby a gel-dye mix is added to the sample which is then loaded onto the chip. The system automatically calculates the ratio of ribosomal bands in total RNA samples using laser induced fluorescence and it provides an image of the ribosomal bands and, in addition, a graph of fluorescence versus time. This can be used to assess the quality of the RNA. The graph shows a number of peaks, the first is a marker and the others correspond to ribosomal RNA. The highest two correspond to the 18 S subunit and the lower to the 5 S and the 28 S. By this analysis, the total RNA extracted from the samples appeared to be of high quality and suitable for use on microarrays.

FIG 3.6.3/3.6.4

SCGT technicians then converted the mRNA to cDNA using reverse transcriptase and an oligo-dT primer, followed by the production of biotin-labelled cRNA. The cRNA was then run again on the agilent to obtain a graphical representation of the quality. It was at this point that problems were encountered. The representation obtained was not what was expected, rather there was a very large peak at about 28 seconds. As we were unable to explain the appearance of this peak we decided not to run the array. I contacted the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow to obtain an Agilent profile of *Drosophila* cRNA which, when subsequently hybridised to a chip, had given good results. As can be seem in figure 3.6.5, the two profiles are very different. Although there is a peak present at 28 seconds in both profiles, in the case of the SHWFGF profile this is smaller than the main "hump". Unfortunately, despite providing fresh RNA samples the same problem was encountered.

FIG 3.6.5

Although, as judged by the total RNA profile provided by the agilent, the starting RNA appeared to be of high quality I decided to try different extraction methods. Therefore I extracted RNA from mutant and wild-type embryos, again using the Sigma kit column-based method and also two other methods, namely Trizol alone, and Trizol followed by a clean-up incorporating the column-based method. In addition I was extremely careful to remove all traces of bleach from the embryos following dechorionation in case carry over of bleach was interfering in downstream reactions. In all cases the total RNA appeared to be of high quality when

run on the agilent but the cRNA from all samples gave the unwanted peak at 28 seconds.

To try to understand what might be happening, I decided to use these probes on a test array. The test array is used to ensure that the cRNA is suitable for use on the much more expensive full *Drosophila* array. The test array contains housekeeping genes with probes to the 5', middle and 3' ends of each gene. A signal value is obtained from each of the three probes for each gene, which indicates the level of expression of that particular gene. The 3'/5' ratio of the signal values for each gene can be used to determine how well the *in vitro* transcription worked. A ratio of less than three suggests that a full length IVT has taken place. The data obtained from the test array are shown in Tables 3.6.1 and 3.6.2.

	Signal (5') Arbitrary units	Signal (M') Arbitrary units	Signal (3') Arbitrary units	Signal (3'/5') Ratio
GAPDH	7.8	21	12.8	1.65
EIF-4A	160.2	880.4	762.9	4.76
ACTIN	88.3	305.6	105	1.91

Table 3.6.1 Test array values for *ato^{1}* SCGT

	Signal (5') Arbitrary units	Signal (M') Arbitrary units	Signal (3') Arbitrary units	Signal (3'/5') Ratio
GAPDH	8.7	11.7	12.8	1.47
EIF-4A	187.1	1139.7	762.6	4.08
ACTIN	111.2	435.5	148.4	1.33

Table 3.6.2 Test array values for *OrR* SCGT

For GAPDH and actin the 3'/5' ratios are satisfactory but the signal values obtained are extremely low. This suggests a fault in the labelling procedure, either because the IVT reaction failed or the initial cDNA synthesis failed. Labelling for microarray experiments is very sensitive to the quality of the primer, which varies between companies and even between batches. In order to test this, I extracted RNA from wild type embryos, employing two different primers to synthesise cDNA, the original one from MWG and a new one from Genset. In addition to synthesising cDNA from *Drosophila,* cDNA was also synthesised from RNA extracted from mouse samples. In the case of the mouse RNA, the Genset primer gave a far superior profile, indicating that there was indeed a problem with the MWG primer. However, this did not solve the problem because the profile obtained from the *Drosophila* sample when using the Genset primer still had the previously observed peak at 28 seconds.

FIG 3.6.6

3.6.4 Microarray experiments carried out at the SHWFGF

Although the SCGT has much Affymetrix and custom array experience, this experience lies wholly in mouse and human arrays. Rather mysteriously, *Drosophila* Affymetrix arrays require their own optimisiation and expertise. SCGT did not have this but the Glasgow facility (SHWFGF) have extensive experience with *Drosophila* Affymetrix microarrays. I decided to make use of this expertise.

In designing the next experiment, I decided to add overexpression analysis to the comparison. I misexpressed *ato* in embryos by *scabrousGAL4* crossed with

UAS-ato to activate *ato* throughout the neuroectoderm. In the overexpression embryos, additional cells express *ato*. Given the ability of *ato* to transdetermine other SOPs, this is expected to result in the upregulation of target genes. The *scabrousGAL4* driver has been used successfully for targeted misexpression in previous microarray studies concerning the nervous system (Egger *et al.*, 2002).

FIG 3.6.7

For these experiments, I stored the embryos in Trizol until I had obtained sufficient numbers for RNA extraction. When I had obtained sufficient embryos, I extracted RNA using Trizol, followed by a clean up using the kit from Qiagen. This approach routinely gave RNA with an A_{260} / A_{280} ratio of more than 1.8. I then transported the samples on ice to SHWFGF. The Agilent profiles obtained for the total RNA were very similar to those generated at SCGT. However, cRNA profiles were different. They were similar to ones previously obtained by SHWFGF for cRNA that had been successfully hybridised to arrays. In the case of the *ato* mutant, there was no peak at 28 seconds, for wild type the peak was evident but smaller than the main "hump" and, for overexpression, the peak was the same size as the "hump". The IVT profiles obtained for all three samples were considered good enough to proceed with the array experiments. The samples were first hybridised to test arrays and the 3'/5' ratios obtained are shown in Table 3.6.3.

	Actin	GAPDH	EIF-4A
ato^{1}	2.13	3.75	4.00
OrR	1.71	2.81	2.81
scaGal4 x UAS ato	2.44	2.26	1.56

Table 3.6.3 3'/5' Ratios for *ato*^{1}, *OrR* and *scaGal4 x UAS ato*, SHWFGF

As most of the ratios were under three, it was likely that full length cRNA had been generated and so the cRNA was hybridised to *Drosophila* genome arrays. The experiment was successful in so far as it yielded a list of the signal values for all genes present on the chip, under the three different conditions. However, rather than repeat the experiment to obtain replicates that could then be analysed, I chose a different approach. This was because I had some concerns about the suitability of using the mutant due to the large number of unfertilised eggs in the mutant embryo collection.

Therefore, for subsequent experiments I decided to use *scabrousGal4 x UAS ato* compared to *scabrousGal4 x UAS nlsGFP* as a control. I chose *UAS nlsGFP,* as its overexpression should not induce any change of expression in genes involved in neurogenesis. I collected embryos from one hour collections of *scabrousGal4 x UAS ato* and *scabrousGal4 x UAS nlsGFP*, aged the embryos for a further 6.5 hours, homogenised the sample and stored it in Trizol at – 80 °C until sufficient embryos had been obtained. I then extracted RNA using Trizol followed by a clean up using the Qiagen column based method. Again, samples were transported on ice to SHWFGF, where, following quality control checks on the RNA and cRNA using the Agilent bioanalyzer, the samples were hybridised to arrays. This experiment was

carried out in duplicate and resulted in a signal value for each of the genes on the chip.

3.7 Data Analysis

I now had two different sets of data, the first consisting of data obtained from the loss-of-function versus wild type versus gain-of- function experiment, i.e., mutant versus wild type versus overexpression, the second consisting of duplicates of the *scaGal4 x UAS ato* versus *scaGal4 x UAS nlsGFP* experiments. All results in the following sections are based on data from the latter experiments.

For robust microarray analysis it is necessary to carry out experiments a minimum of three times. Time did not allow this and so it is difficult to draw concrete conclusions from the data obtained. Nonetheless, it is informative in some ways and can certainly be built upon in the future.

3.7.1 Description of steps taken to manipulate data

Scanning records the pixel- by -pixel intensity for each probe and a raw image is produced of different fluorescent spots. Based on the probe sets for each gene, array software converts the pixel-by-pixel data to give a signal value for each of the genes represented on the array. As described previously, for each gene there are multiple repeats and corresponding mismatches. Ideally gene X, would hybridise avidly to its perfect match but should not hybridise to its corresponding mismatch. Any fluorescence due to the mismatch is taken as non-specific hybridisation and it is subtracted from the fluorescence of the perfect match. On *Drosophila* chips there are 14 probe pairs per gene and the signal value for a particular probe set is defined as

the mean of all the (PM-MM) differences. The signal value is taken as proportional to the actual amount of RNA of the corresponding gene in the sample. The signal value, coupled with other factors such as variation between different probes for a particular gene, is used to give an Absolute Call. This is where the transmit for each gene is termed Present (P), Marginal (M) or Absent (A). This analysis is carried out on each individual array. Table 3.7.1 shows the number of genes called Present or Marginal for each of the four arrays. The total number of genes called present or marginal on at least one array is 7,863 (out of more than 13,500).

	Ato 1	Ato 2	GFP 1	GFP 2
Total no. of P/M genes	7451	6890	6727	6805

Table 3.7.1 Total number of present/marginal genes on each individual array.

As the Affymetrix system uses only one fluorochrome per chip, two arrays are required to compare a sample and control. In order to do this, it is necessary to normalise and scale each data set relative to each other to account for technical differences between experiments, such as the efficiency of probe labelling and hybridisation. Normalisation was carried out by the SHWFGF. Obviously, scaling is necessary not only to compare sample and control but also to compare replicates of sample and control. Any technical variability experienced during the experiment will have an identical effect on the signal values of non-differentially expressed genes as on those that are expressed differently between experiments. As most genes present on the array belong to the former category, standard practice is to use either the median signal or the total signal for all detected genes as a correction factor.

In this case, the basic scaling factor used is the sum of the signals for all genes which give a present or marginal call. Scaled data for each individual array is obtained by dividing the signal for each gene by the sum of the P/M signals. As this number is very small I multiplied it by 10^5 to obtain more manageable figures. This scaled data is used in subsequent analysis.

Once scaling has been carried out on each individual array, values from the replicates were averaged. This results in an average for the expression of each particular gene in the sample and in the control. This means that it is possible to compare the expression level of any gene in the sample to the expression level of the same gene in the control. This involves calculating the ratio of the mean of *(UAS ato)* to the mean of *(UAS nlsGFP)*. It is standard practice to use a log scale to represent this ratio. Therefore, I calculated the $\log_2$ *(UAS ato)/(UAS nlsGFP)*. Using $\log_2$ values means that a change of 1 log represents a two-fold change in the data.

A subset of the genes are up- or down-regulated in *UAS ato* as compared to *UAS nlsGFP*. It is then necessary to determine if the change in expression is significant or if it is the result of noise. A t-test can be used to determine whether the expression of a particular gene is significantly different between control and sample. Clearly, the greater the number of replicates the better. It is then possible to sort the list of expressed genes in order of the best (i.e., lowest) p values. Although, instinctively one might assume that it is best to order the expressed genes by those that have the highest $\log_2$ ratio (i.e., those that show the greatest difference in expression), this is not necessarily the case as high log ratios may be the result of noise; it is consequently essential to take the p values into account. This is especially

relevant here as many genes downstream of *ato*, will not be very highly expressed and so may have quite low fold changes and will be most subject to noise.

3.7.2 Effectiveness of Normalisation/Scaling

The global data can be represented graphically in a number of ways. A log - log plot of the scaled signal for each gene in the sample versus the signal of each gene in the control is shown in figure 3.7.1. An MA plot is used to show the intensity dependent ratio of raw microarray data. It illustrates intensity specific effects by plotting the log ratio of each gene as a function of the square root of the product of the individual intensities. If normalisation/scaling has been carried out successfully, the plot should show a straight line as is seen in fig 3.7.2.

FIG 3.7.1

FIG 3.7.2

As the purpose of scaling is to minimise signal differences generated by technical variation, the sum of the scaled values should be similar across experiments. The values obtained for each array are shown in table 3.7.2.

	Ato 1	Ato 2	GFP 1	GFP 2
∑ unscaled signal values	3,649,920	2,304,694	2,165,740	2,519,751
∑ scaled signal values	17,072	15,599	15,728	15,487

Table 3.7.2 Comparison of the ∑ unscaled and ∑ scaled signal values.

Examination of the sum of the unscaled signal values obtained for each array shows that the first *scaGal4 x UAS ato* array has the highest sum of signal values and differs significantly from the other three arrays. This is probably due to technical variation. Following scaling, however, this difference is less pronounced, showing that the scaling is successful in minimising technical differences between experiments. As mentioned previously, it is possible to either use the sum of the Present and Marginal signals or the median of the Present and Marginal signals as a scaling factor. There appears to be no particular advantage to using one over the other and, in this case, the sum of the signals was used. Scaling appeared to be successful and it was now possible to examine the values obtained in more detail.

3.7.3 Values obtained for a number of sample genes

I obtained the fold change and a p value for each of the genes that were given an Absolute Call of Present or Marginal on the array. I then examined these values for a number of genes. The first gene I considered was *ato*. Clearly there should be differences in the expression level of *ato* in *scaGal4, UAS ato* when compared with *scaGal4, UAS nlsGFP,* and this indeed was the case (Table 3.7.3). When using the sum of the P/M signals as a scaling factor, the $\log_2$ ratio of the signals ($*10^5$) is

2.623. As this is a log scale, this means that there is a six fold change in expression between the sample and control and the p value is 0.011. When unscaled signal values are used, the $\log_2$ ratio of the signals is 2.971 (i.e., fold change is almost 8) and the p value is 0.128. The importance of scaling is clearly illustrated here, as unscaled values give a much higher p value. (The lower the p value, the greater the likelihood that any change is significant). The unscaled p value represents not only the true change in gene expression but also any change due to technical differences between arrays.

As a control, I considered a gene that should not show a difference in expression levels. This gene was *histone3A,* (Tabel 3.7.4). This gave a $\log_2$ ratio of 0.057, which is a fold change of one, and a p value of 0.431 for the scaled values. This is a low fold change and a high p value, indicating only a slight possibility that the change is significant. Stathopoulos *et al.* (2002) state that a fold change of less than one is what one would expect for a uniformly expressed gene. An increased number of replicates would probably reduce the fold change even further. I also looked at the values obtained for *twist,* a bHLH transcription factor expressed in muscle (Table 3.7.5). The expression of this gene should not change in the sample compared to the control as it is not involved in neurogenesis. It is however, involved in myogenesis, another developmentally regulated process. The fold change obtained was less than one and the p value 0.556. The combination of a low fold change and a high p-value suggests that any change is not due to a true change in gene expression, i.e., there are no expression changes in genes involved in other developmentally regulated processes.

	Ato 1	Ato 2	GFP 1	GFP 2	Ratio of means	Log_2 (ratio)	t-test
Unscaled signal value	1905.8	1325.6	224.2	187.8	7.84	2.971	0.128
Scaled signal value*10^5	52.215	57.517	10.352	7.453	6.16	2.623	0.011

Table 3.7.3 Comparison of values obtained for log_2 (ratio) and t-test for *ato* when using scaled and unscaled signal values.

	Ato 1	Ato 2	GFP 1	GFP 2	Ratio of means	Log_2 (ratio)	t-test
Unscaled Signal value	6930.6	4617.6	4184	4589.9	1.32	0.401	0.438
Scaled Signal value*10^5	189.884	200.356	193.19	182.157	1.04	0.057	0.431

Table 3.7.4 Comparison of values obtained for log_2 (ratio) and t-test for *histone3A* when using scaled and unscaled signal values.

	Ato 1	Ato 2	GFP 1	GFP 2	Ratio of means	Log_2 (ratio)	t-test
Unscaled signal value	172.2	184.5	191.6	168.9	0.99	-0.015	0.900
Scaled Signal value*10^5	4.718	8.005	8.847	6.703	0.82	-0.289	0.556

Table 3.7.5 Comparison of values obtained for log_2 (ratio) and t-test for *twist* when using scaled and unscaled signal values.

3.7.4 Genes involved in neurogenesis

The number of genes called present on at least one array is 7,863 (out of more than 13,500). Present genes, ranked on the basis of p values, were plotted in the order of decreasing significance of differential expression (figure 3.7.3/3.7.4).

FIG 3.7.3/3.7.4

The data set was searched for genes that are known to be involved in neurogenesis to determine the fold change and p value. Table 3.7.6 shows the fold change, p value and position on the rankings for some of the known genes.

ato is significantly different in both datasets. Out of the 7,863 P/M genes that are present on at least one array, *ato* is reasonably high in the fold change and p value rankings, 60 for the fold change and 71 for the p value. It is not first however, and this illustrates the problems inherent in data analysis. Higher ranking genes seem unlikely to be involved in neurogenesis. Many have a much higher fold change than *ato* or are expressed at much higher levels than *ato*. This seems unlikely for target genes. In addition, although *ato* has a reasonable fold change (six fold), the data is too noisy to give a lower p value as there are only two replicates. Further replicates should eliminate some of the noise and give a lower p value. As it stands, the p value of 0.011 is probably too high to be automatically chosen as a true positive.

I also searched for some of the "generic" target genes such as *E(spl), Brd, sca, ase* and *sens*. I would expect the fold change to be very small as *ato* is not the only regulator of these genes. This was borne out by the values obtained. *senseless* is not present on the array and apparently has a bad probe set. This is also the case for *cut*. I also looked at the values for the other proneural genes and for *da*. Although *amos* shows a reasonable fold change it has a high p value indicating there is a high likelihood that this is due to chance. The other proneural genes and *da* show a low

fold change and a high p value demonstrating they are not differentially regulated. As mentioned previously, there are some known differentially regulated target genes of *ato,* such as *rhomboid* and *TAKR86C*. *rho* shows poor detection suggesting the possibility of a bad probe set; *TAKR86C* is present on one array and not detected on the others suggesting that the present call in one case may be due to technical variation. The inability to detect a change in the expression levels of *TAKR86C,* although it is a known target gene of *ato,* illustrates the limitations to the approach of using whole embryos. Microarray analysis on whole embryos may not be sensitive enough to detect fold changes for genes expressed at a low level.

Gene	**Mean (ato)**	**Mean (GFP)**	**Ratio**	**Log_2 (ratio)**	**t-test**	**Rank (ratio)**	**Rank (t-test)**
atonal	54.87	8.90	6.16	2.62	0.011	59	71
amos	4.30	1.42	3.02	1.59	0.501	150	4559
scute	6.04	6.76	0.89	-0.16	0.618	5820	5411
achaete	1.63	1.40	1.17	0.22	0.709	2092	6032
l'scute	12.70	14.50	0.88	-0.19	0.705	6053	6002
da	30.65	31.57	0.97	-0.04	0.874	4669	7077
asense	7.25	6.04	1.20	0.26	0.241	1801	2196
cato	2.93	2.81	1.04	0.06	0.815	3543	6713
scabrous	6.64	5.70	1.16	0.22	0.575	2102	5081
TAKR86C	0.07	0.36	0.20	-2.30	0.518	7857	4681
rhomboid	2.01	1.42	1.42	0.50	0.366	814	3391
Delta	15.63	20.10	0.78	-0.36	0.011	6985	68
E(spl)	39.35	26.99	1.46	0.54	0.153	524	1287
Bearded	105.16	80.69	1.30	0.38	0.161	1198	1376

Table 3.7.6 Fold change, p value and position on the rankings for some genes known to be involved in neurogenesis.

When genes are ranked either by fold change or p values, it is not immediately clear even in the case of *ato,* that there is a true change in expression. The raw data i.e. the signal values indicate that this is the case and *ato* has a reasonable fold change but is ranked only 71st in the p value list suggesting, that in the absence of more than two replicates, the signal to noise ratio is too high to identify *ato* automatically as a true positive.

Many of the genes that have high ratios are serine proteases, muscle structural proteins, cuticle proteins and immune response proteins. However, these genes tend to have high p values and so show high variation between arrays. In some cases, these appear to be larval or adult proteins. This suggests that one of the samples may have been contaminated by larvae or part of a dead fly. This illustrates the need for extreme care when collecting embryos for RNA extraction and subsequent microarray analysis.

3.7.5 Consideration of the top ranking genes

In order to assess the best ranked genes more generally, I examined the fold change and p values in parallel. Egger *et al.*, (2002) considered transcripts as being differentially expressed only if they showed both a fold change of ≥ 1.5 or < 1.5 and a significance value of $p < 0.01$. The cut off point for fold changes is not fixed and depends on each individual data set. Arbeitman *et al.*, (2002) for instance when searching for differentially expressed genes during the life cycle of *Drosophila,* used a four fold increase as the cut off point for differentially expressed genes while Stathopoulos *et al.*, (2002) used a three fold increase as their criterion.

In this case, the fold change obtained for *ato* itself is six. Given my expectation that *ato* regulated genes will be expressed at a low level I considered all the genes that had a fold change greater than 0 i.e., showed some change in expression. This includes those that are both up and downregulated in the sample when compared to the control. As it seems more likely that overexpression of *ato* will cause an increase

in the expression level of target genes, I concentrated on those genes that were upregulated in the sample compared to the control. A total of 4,103 genes were in this category and I ranked them on the basis of p values and took a cut off p value of 0.05. This is a high p value but because the experiment was only carried out in duplicate, the p values will be higher across the board (including *ato).* More replicates will decrease the p value. Of these 4.134 genes that showed some change in gene expression, 200 had a p value of less than 0.05. *ato* was number 38 on this list. When those genes that are called absent in at least three arrays are removed, the total number of genes that show an increase in expression and have a p value > 0.05 is 190. *ato* now ranks at number 33. The remaining genes fit into the following categories (after Egger *et al.*, 2002). Table 3.7.7.

Molecular Function	Number of transcripts
Nucleic acid binding	6
DNA binding	11
Transcription Factor	8
RNA binding	4
Translation factor	2
Ribonucleoprotein	1
Cell cycle regulator	3
Chaperone	1
Motor Protein	3
Defense/immunity protein	1
Enzyme	29
Kinase/phosphatase	6
Apoptosis regulator	-
Signal transducer	3
Cell adhesion	4
Structural protein	5
Transporter	4
Ligand binding or carrier	8
Antioxidant	-
Tumor suppressor	-
Other	12
Function unknown	72
Total	190

Table 3.7.7 Differential gene expression in functional classes following *ato* misexpression

Although this data is very preliminary and must be repeated before any definite conclusions can be drawn, there are present a number of genes that play a role in neurogenesis. These are discussed below.

Given that *ato* acts as a key regulator of neurogenesis, it appears likely that *ato* may control several other transcription factors. These transcription factors could then in turn regulate expression of downstream target genes. *ato* appears to regulate seven transcription factors at least two of which have been shown to be involved in embryonic nervous system development. *tramtrack (ttk)* is implicated in SOP

formation. In external sense organ SOPs *ttk* is inhibited by *phyllopod,* which allows formation of an external sense organ (Pi *et al.*, 2001). Another highly ranked transcription factor is *nervy (nvy). nvy* encodes a nuclear protein that is expressed in embryonic neuroblasts and SOPs. Wildonger and Mann presented an abstract at the *A. Dros. Res. Conf 41 (2000),* suggesting that *nvy* plays a role in neuroblast identity and SOP formation. In addition there are six genes involved in DNA binding one of which is *klumpfuss (klu). klu* has transcription factor activity and has been implicated in SOP formation. Kaspar and Klein presented an abstract at the *A. Dros. Res. Conf 43 (2002),* where they investigated the role of *klu* in bristle and leg development. *klu* mutants possess a reduced level of SOPs and when *klu* is overexpressed more SOPs are formed. In addition *klu* may be a target of *amos* (T. Klein, pers. comm.).

ato misexpression is also likely to affect genes that encode proteins involved in cell-cell signalling. Three kinases and three phosphatases show a change in expression between sample and control. One of the phosphatases is *brahma,* which is involved in chromatin-mediated maintenance of transcription. *breathless* encodes a fibroblast growth factor (FGF) receptor and shows an increase in transcript abundance. Other differentially expressed genes involved in signalling pathways include *argos-* a negative regulator of EGF receptor signalling, which is involved in the formation of chordotonal organs (zur Lage *et al.*, 1997), and *loco,* a regulator of G protein coupled signalling that is expressed in the peripheral nervous system. Indeed, the Posakony lab (UCSD), have recently identified *loco* as a downstream target of *sc.* The presence of these genes is encouraging and suggests that the approach taken is capable of identifying genes involved in neurogenesis.

3.8 Discussion

The key to having full confidence in microarray experiments is replication. The greater the number of experiments the more confidence one can place in a result. Unfortunately, lack of time prevented me from carrying out the *scabrousGal4 x UAS ato* compared to *scabrousGal4 x UAS nlsGPP* experiment more than twice. In order to have full confidence in the results the experiment should be repeated a minimum of three times. Therefore, the results presented here are preliminary. As experiments of this nature are usually at a minimum carried out in triplicate, I was unable to draw any concrete conclusions as to which genes may be putative downstream targets of *ato,* but I was able to gain an understanding of the microarray process and how best to approach the problem of identifying downstream target genes. In addition, despite the below optimum number of replicates it appears that there is some enrichment of interesting genes in the rankings. One possible approach would be to carry out RNA *in situs* on genes that show up as being differentially expressed to determine their expression pattern. Although most will probably be false positives, it appears likely that some will be true downstream targets. However, the better option is to repeat the microarry analysis at least once more to obtain a more accurate gene expression profile. Preliminary data indicated that it should be possible to identify at least relatively highly expressed target genes. I attempted to integrate the mutant versus wild-type versus overexpression data, but this decreased the number of significant genes suggesting that the conditions and lines used in these experiments were not compatible with the *scaGal4 x UAS ato* versus *scaGal4 x UAS nlsGFP.*

3.8.1 Future approaches

As outlined above, my original experiments focused on mutant versus wild-type versus overexpression and subsequently solely on overexpression. However this approach concentrated only on the proneural gene *ato,* and I am aware that the data generated (even if more replicates were to be carried out), will not in itself directly address the question of proneural gene functional specificity i.e. what distinguishes the proneural genes. I have been given the opportunity to follow up the work outlined in this chapter and in addition to determining the specific downstream targets of *ato,* I propose to compare directly the expression changes induced by misexpression of *sc, ato* and *amos.*

My reasons for proposing this are:

(i) These comparisons will identify specifically those target genes that are differentially regulated by the different proneural genes rather than co-regulated target genes.

(ii) Subsequent analysis will be simplified, since only a subset of proneural target genes will be identified.

(iii) Transcriptome comparisons can be better controlled experimentally because genetic background effects can be minimised.

(iv) Any artificial changes induced by the experimental system (e.g. presence of yeast Gal4) will not be uncovered in the comparison.

I propose to again use the *scaGAL4* driver for targeted misexpression in the neuroectoderm. I plan to do a four way comparison of *scaGAL4* crossed with *UAS-sc* or *UAS-ato* or *UAS-amos* versus *sca-GAL4 x UAS-nlsGFP.* My aim is to perform

this experiment in quadruplicate. This, I hope, will identify the downstream target genes of the proneural proteins. If this is the case, it will then be necessary to validate the data obtained from the microarray experiments. This validation is essential, as micorarray analysis will implicate possibly hundreds of genes in the mode of action of the proneural proteins. I am aware that it is the subsequent analysis that will effectively establish with certainty which genes are the direct downstream targets and, as such, contribute most to solving the question of proneural protein specificity. I plan to carry out *in situ* hybridisation to determine the expression patterns of putative targets. As misexpression may potentially activate expression of genes that are not involved in neurogenesis, validation by examination of the expression pattern in wild type embryos will be vital. I would hope to find clearly defined expression patterns that overlap with expression of the proneural gene, suggesting that the proneural protein is capable of direct regulation of the putative target. I would also be interested in carrying out *in situ* analysis of the targets in proneural mutant and overexpression embryos to determine if there is a change in their expression pattern. This work should validate the microarray data by determining if the putative targets show localized expression patterns in the sense organ precursor cells in a similar time scale as compared to expression of the proneural genes.

In addition, the role of the most promising subset of identified genes in sense organ precursor development, could be assessed by disrupting their function through injection of double stranded RNA (RNAi). The selection criteria will be based on the expression pattern identified by the *in situ* experiments and/or on putative gene function. Furlong *et al.*, (2001) used RNAi when studying the role of the bHLH

gene, *twist,* in mesoderm development. I shall inject dsRNA in embryos and then assess sense organ precursor development by immunohistochemistry and confocal microscopy. I anticipate likely phenotypes to be a disruption of SOP number, pattern or identity, or disruption of neuronal differentiation.

In addition, I would like to utilise bioinformatic methods to search for *cis*-regulatory elements in the target genes. I aim to analyse the different sets of differentially regulated genes to look for a) proneural protein binding sites and, b) other shared *cis*-regulatory sites, which may be bound by transcription cofactors in concert with the different proneural proteins. Stathopoulos *et al.* (2003) used a *cis*-regulatory bioinformatic package to identify all possible 7-11 base pair sequence motifs shared between the target genes identified for Dorsal. In addition to identifying known motifs, three novel shared motifs were also recognised. This demonstrates clearly the tremendous amount of information available in the microarray data. The power inherent in comparing the enhancer regions of many targets is obvious and should potentially reveal a great deal about the binding sites present in these genes.

Identification of proneural binding sites in the target genes will help elucidate how proneural proteins regulate these genes. In addition, identification of both known and novel motifs for other transcription factors will implicate putative binding partners which will open the way to pursue the "cofactor hypothesis". This is based on the belief that specificity is not due to different DNA binding between the proneural proteins but rather to protein-protein interactions with subtype specific cofactors (Brunet and Ghysen, 1999, Chan and Jan 1999).

3.8.2 Limitations of the whole embryo approach

Although results obtained from the *scaGal4 x UAS ato* compared with *scaGal4 x UAS nlsGFP* are preliminary, to some extent they illustrate the viability of the approach. By this I mean, the fact that known neurogenic genes can be assayed in these experiments. Although encouraging, this comes with the caveat that it is not easy to identify these genes on the basis of the fold change and p values. Although more replicates would assist in identifying these genes this may be a continuing problem. All microarray experiments experience a degree of noise. Data obtained from arrays is only of use if the fold change due to specific hybridisation is sufficiently high when compared to that obtained by non-specific hybridisation i.e. the signal to noise ratio. In the case of *ato,* the fold change due to *ato* regulation appears to be low when compared to the noise and so it may be difficult to detect differentially expressed genes. In whole embryos a high proportion of the cells are not relevant to SOP formation and hence the RNA from differentially expressed genes is likely to consist of only a small proportion of the total mRNA pool. This means it is possible that the signal to noise ratio may be too low to detect a meaningful change in gene expression. *TAKR56C* is an illuminating case: it just cannot be detected on the microarrays. Although it is a target of Ato, it is not sufficiently abundant at the whole embryo total RNA level. Therefore, I explored the possibility of using cell sorting to obtain pure populations of neural precursor cells which could be used in micorarray analysis. This is discussed further in chapter 4.

3.9 Conclusions

This chapter has focused on my efforts to identify downstream targets of the proneural gene *ato,* using misexpression in whole embryos. Despite severe technical difficulties, I succeeded in carrying out duplicate experiments for *scaGal4 x UAS ato* compared to *scaGal4 x UAS nlsGFP.* More replicates are required before candidate targets can be confidently identified but this pilot study has proved the feasibility of applying microarray analysis to the challenge of identifying proneural target genes. I became concerned that analysis of whole embryos might not be able to identify more weakly expressed target genes and therefore, as outlined in chapter four, investigated the possibility of obtaining pure populations of neural precursor cells.

Future plans include comparing directly the expression changes induced by misexpression of *sc, ato* and *amos.* These genes play a profound role in neurogenesis and it is hoped that this work will contribute significantly to what is known about their mechanism of specificity. The combined approach of microarray analysis followed by *in situ* hybridisation will identify many of the genes that are differentially regulated by this closely related family of transcription factors. This will give an indication of the kind of genes that require activation by the proneural genes to function in neurogenesis. The RNAi studies will then elucidate the function of a subset of these. The computational methods of searching for enhancer elements in these genes will be highly effective in identifying conserved motifs and hence provide insight into how these target genes are regulated by different proneural proteins via both DNA binding site requirements and possible cofactor involvement.

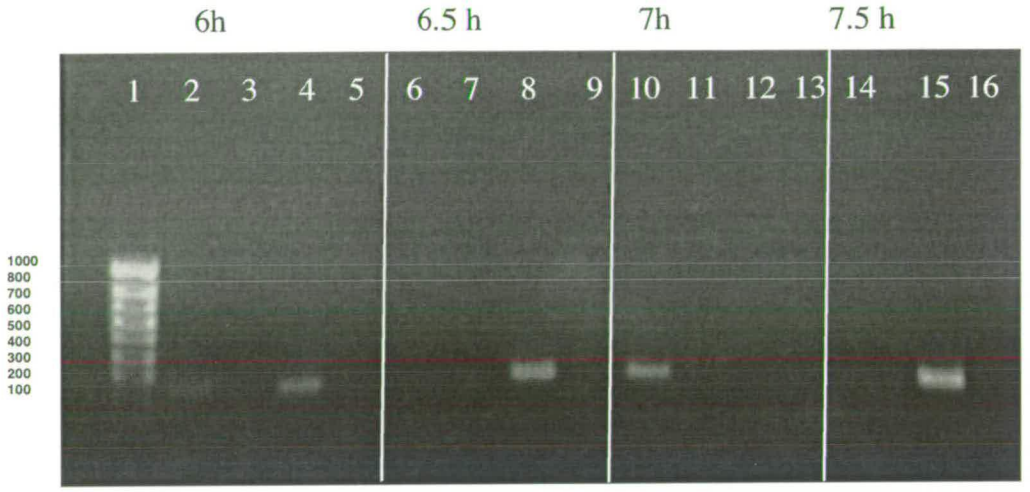


Fig 3.6.1 Gel showing expression of *ato* in *ato*^{1} and *OrR* flies at 6.0, 6.5, 7.0 and 7.5 h after egg laying. Lane 1. Marker. Lane 2, *ato*^{1} 6.0 h, plus RT. Lane 3, *ato*^{1} 6.0 h, minus RT. Lane 4 *OrR* 6.0 h, plus RT. Lane 5 *ato*^{1} 6.5 h, minus RT. Lane 6 *ato*^{1} 6.5 h, plus RT. Lane 7,, *ato*^{1} 6.5 h, minus RT. Lane 8 *OrR* 6.5 hr, plus RT. Lane 9 *OrR* 6.5 hr, minus RT. Lane 10. *ato*^{1} 7.0 hr, plus RT. Lane 11, *ato*^{1} 7.0 h, minus RT. Lane 12 *OrR* 7.0 h, plus RT. Lane 13 *OrR* 7.0 h, minus RT. Lane 14, blank. Lane 15 *ato*^{1} 7.5 h, plus RT. Lane 16 , *ato*^{1} 7.5 h, minus RT.

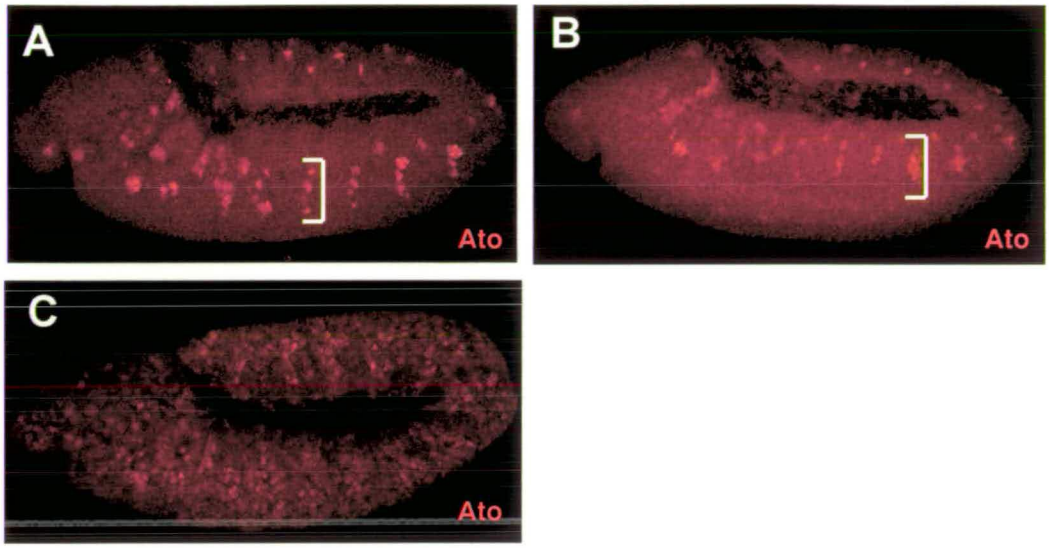


FIG 3.6.2 *ato* expression in wild-type, mutant and overexpression embryos 6.5-7.5 h after egg laying. Confocal image of embryos 6.5-7.5 AEL (stage 11) stained with antibody to detect Ato protein in red. (A) Wild type *OrR*, expression is resolved to four cells per hemisegment (bracket) (B) *ato*^{1} *ato* expression is not resolved to four cells per herimsegment (bracket) (C) *scaGal4 x UAS ato*.. Overexpression turns more cells throughout the neuroectoderm into *ato* expressing cells.

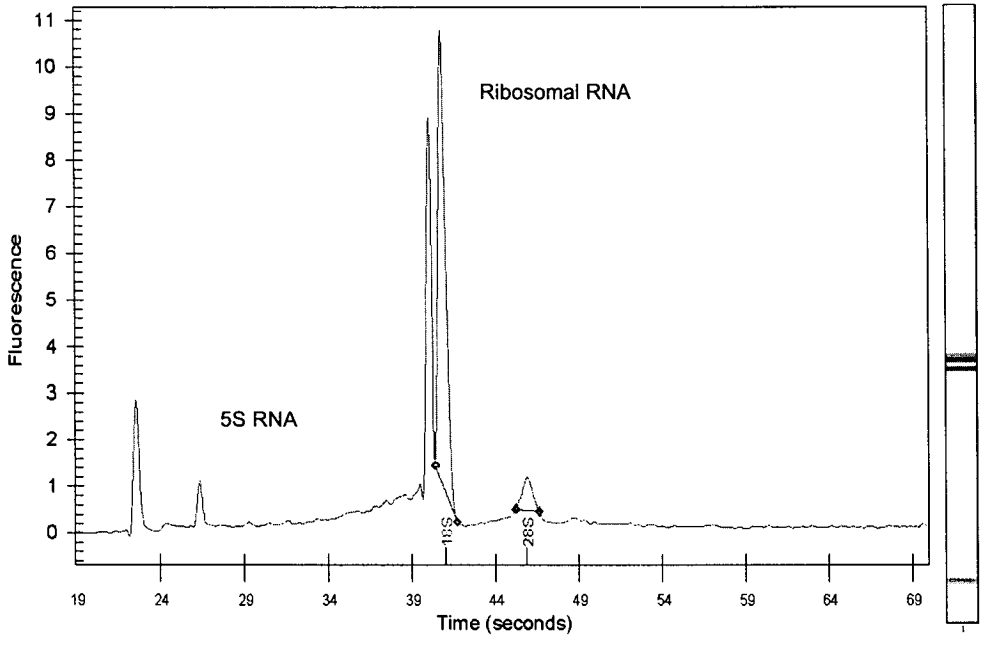


FIG 3.6.4 Agilent profile of Total Drosophila RNA- graphical representation. Fluorescence versus time.

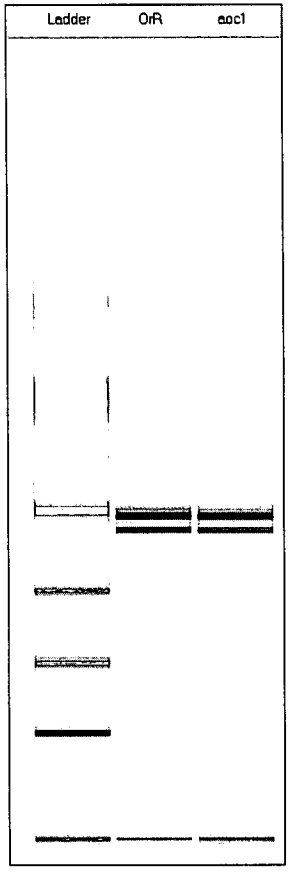
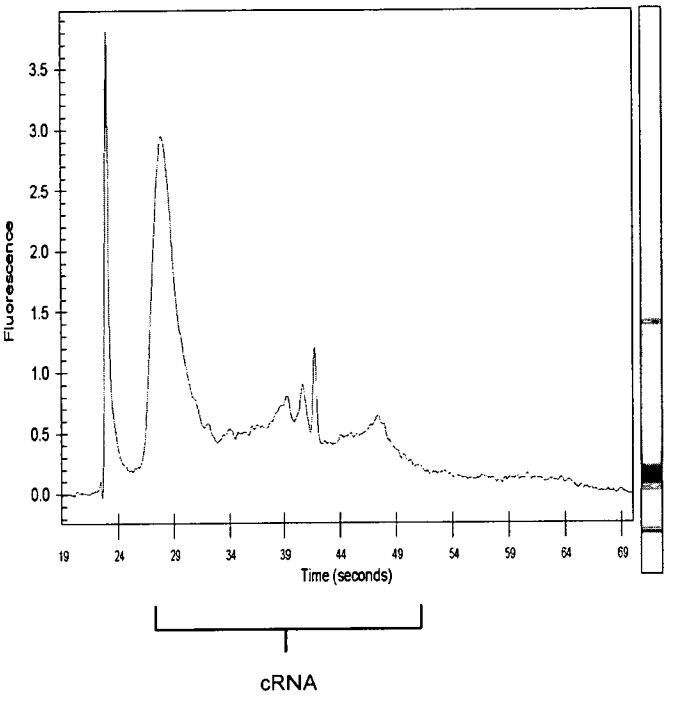


FIG 3.6.3 Agilent profile of total RNA. Lane 1, Ladder. Lane 2 *OrR* Lane 3 *ato*^{1}.

A



B

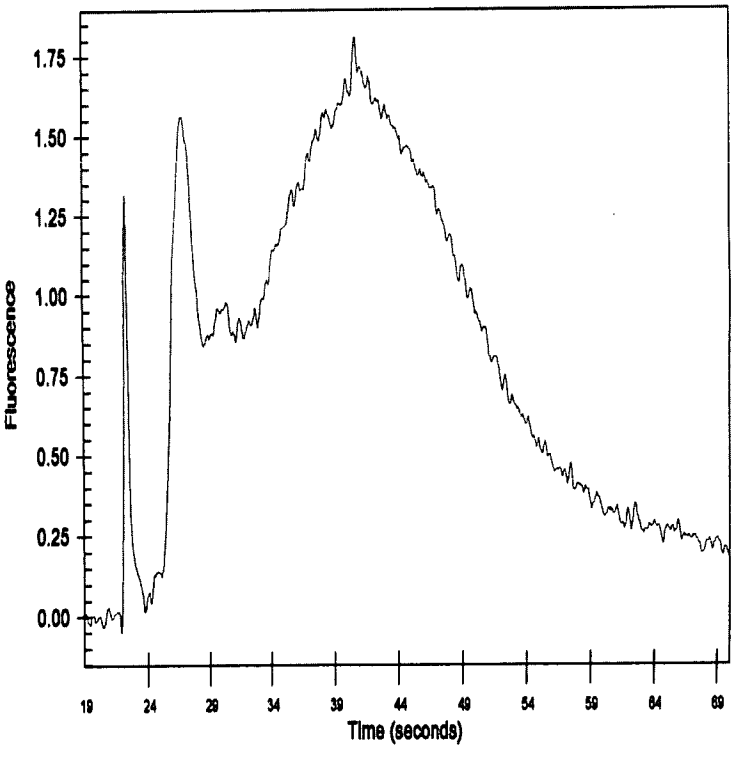


FIG 3.6.5 Agilent profiles of labelled cRNA (A) SCGT (B) SHWFGF. Fluorescence versus time. Peak visible at 28 s in both profiles but main "hump" absent in SCGT profile.

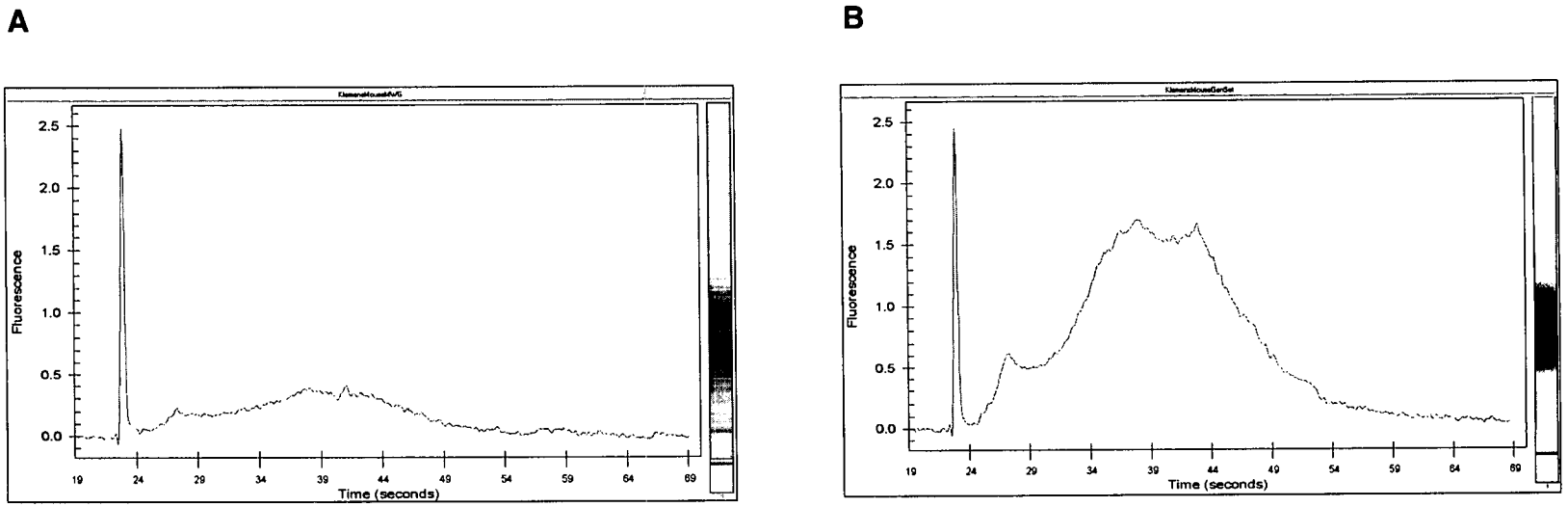


FIG 3.6.6 Comparison of mouse cRNA using primer from MWG and primer from Genset.
(A) mouse sample MWG primer. (B) mouse sample Genset primer

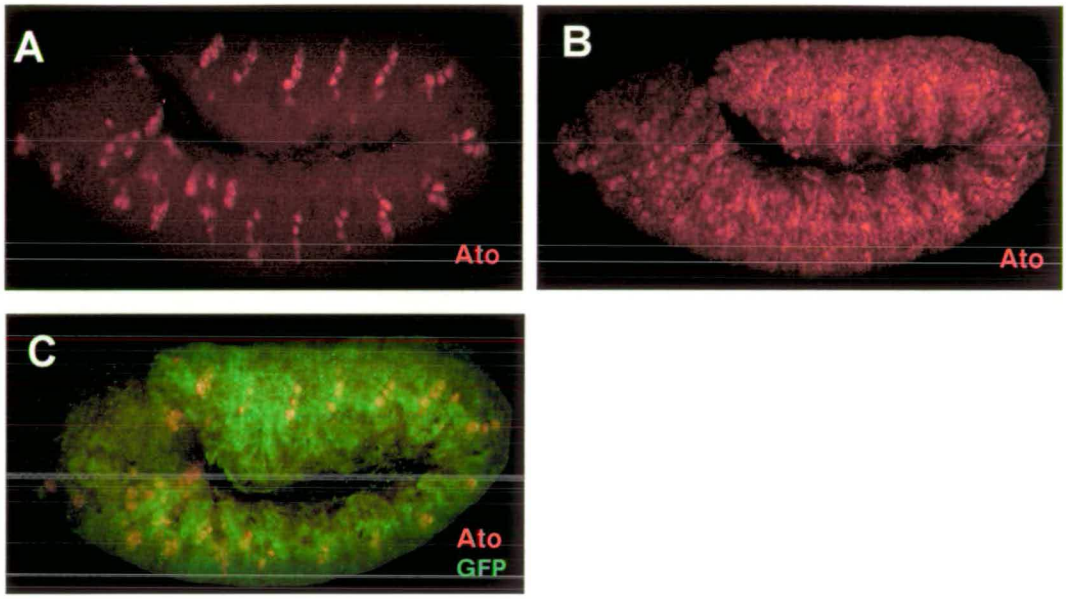


FIG 3.6.7 Misexpression of *UAS ato* and *UAS nlsGFP* using *scaGal4* , 6.5-7.5 h after egg laying. (A,B) Confocal images of embryos 6.5-7.5 h AEL (stage 11) stained with antibodies to detect Ato protein in red. (A) Wild type (*OrR*). *ato* expression is resolved to four cells per hemisegment. (B) *scaGal4 x UAS ato* , overexpression turns more cells into *ato* expressing cells. (C) Confocal image of embryo 6.5-7.5 h AEL (stage 11) stained with antibodies to Ato protein in red and GFP protein in green. *scaGal4 x UAS nlsGFP*, overexpression turns more cells into GFP expressing cells.

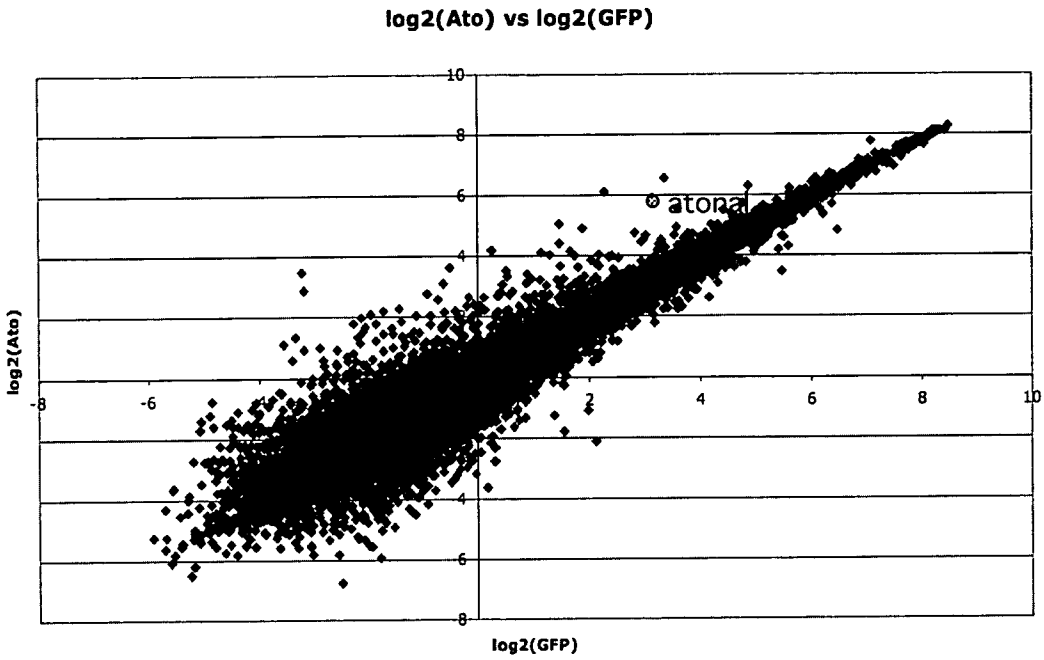


FIG 3.7.1 Log-log plot of *scaGal4 x UAS ato* and *scaGal4 x UAS nlsGFP* data. Log_2 (Ato) versus Log_2 (GFP).

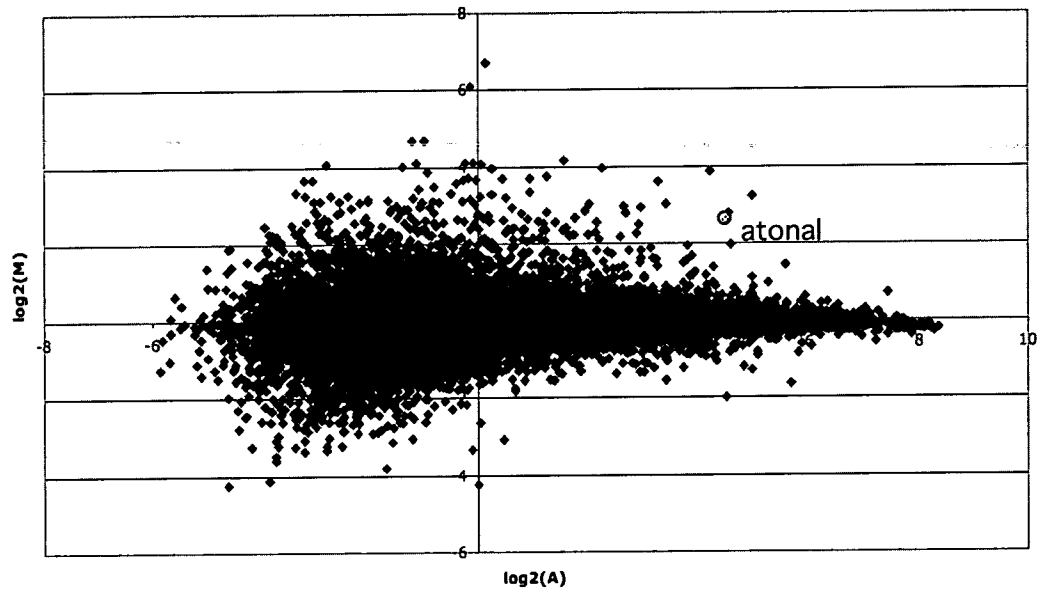


FIG 3.7.2 MA plot of *scaGal4 x UAS ato* and *scaGal4 x UAS nlsGFP* data. Log_2 (M) versus log_2 (A). M = Ato/GFP A=SQR(Ato*GFP).

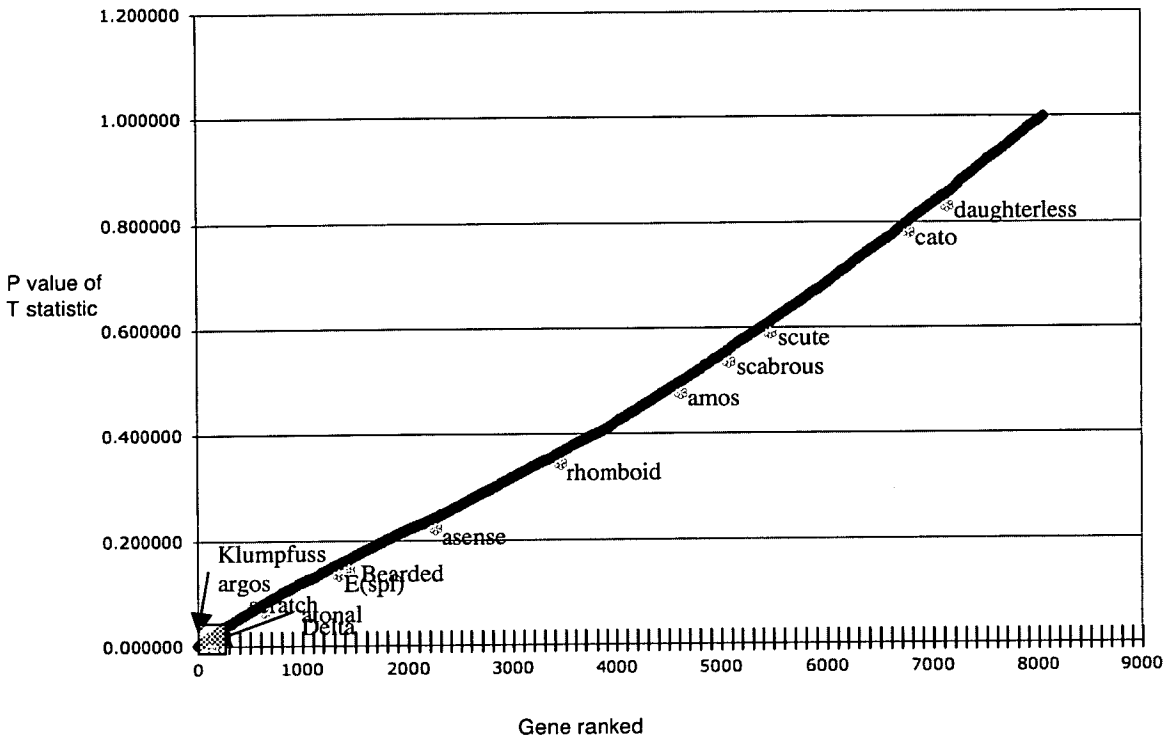


FIG 3.7.3 'Present' genes in order of decreasing significance of differential expression.
P value of T statistic versus gene ranked.

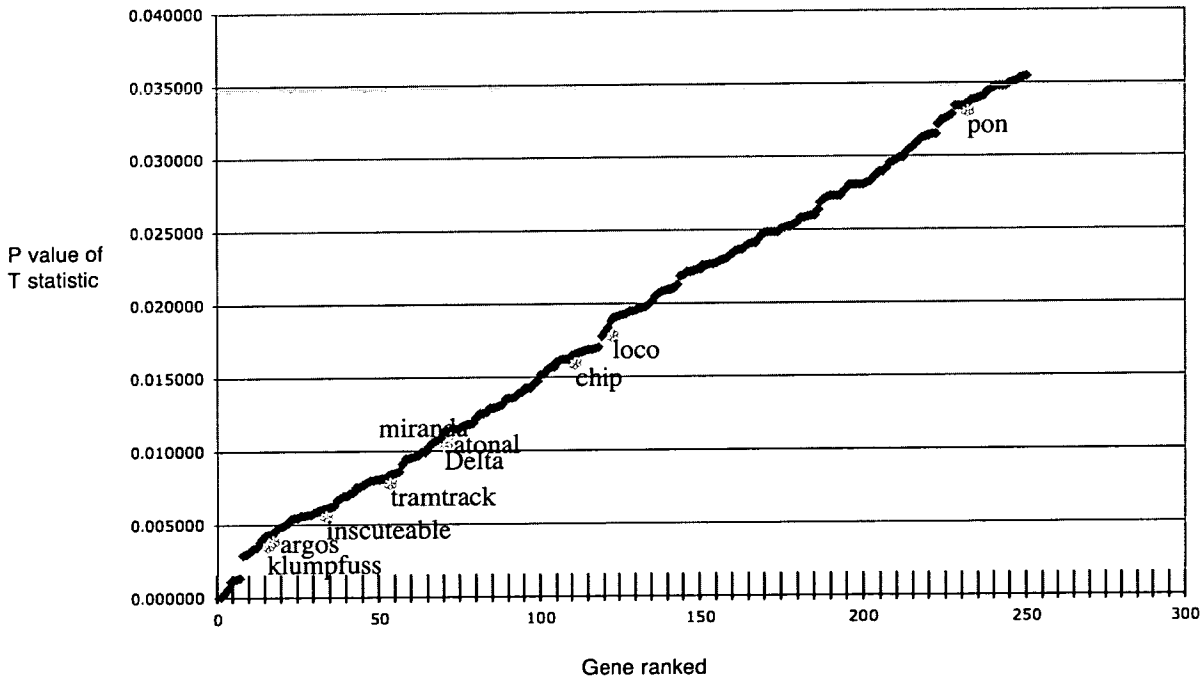


FIG 3.7.4 Top 250 genes in order of decreasing significance of differential expression.
(Square region on 3.7.3). P Value of T-statistic versus gene ranked.

4

Exploring cell sorting as a way to compare sense organ precursor transcriptomes.

4.1 Aims of this Chapter

The overall aim of this chapter is to develop a system to isolate specific subpopulations of neural precursor cells.

The approach taken was as follows,

1. Obtaining or constructing fly lines that express green fluorescent protein (GFP) in the neural precursor cells of the imaginal discs and/or embryos.
2. Isolation of these cells by virtue of the GFP expression using cell dissociation and fluorescent activated cell sorting (FACS) analysis.

The long-term goal is to use these cells in microarray analysis. As each proneural gene is responsible for a different subset of neural precursor cells, and hence for different types of sense organs, comparison of the transcriptomes of these subpopulations should result in identification of specific downstream target genes for each proneural gene. It is these differential downstream target genes that allow specific SOPs to take on different identities and functions. Alternatively the transcriptional profiles of neural precursor cells for each subtype of sense organ could be compared with non-neural cells.

4.2 Microarray analysis on sorted cells versus whole embryos

Taking whole embryos is subject to the limitation that there is a high proportion of cells which are not relevant to sense organ precursor formation present in the sample. Only a very small proportion of the cells present in each embryo are expressing proneural genes in the PNS and hence the RNA from differentially expressed genes is likely to consist of only a small proportion of RNA contributing to the total RNA pool. This means it is possible that the ratio of signal to noise may be too low to detect a meaningful change in gene expression. This is a concern even when an overexpression system is used to turn more cells into the cells of interest. The development of a system to obtain a pure population of neural precursor cells would be a significant advance on the method of isolating RNA from whole embryos. This would greatly increase the probability of obtaining direct downstream targets specific to each population of neural precursor cells and hence each proneural gene.

There are, however, some disadvantages to this approach. These include the possibility that the experimental manipulation involved may alter the transcriptional profile of the cells. Another disadvantage from a practical point of view is the scarcity of material. Means of overcoming these problems will be discussed in later sections.

4.3 Imaginal disc cells versus embryonic cells

Proneural genes are expressed in the embryo and in imaginal discs. An important question is whether embryos or imaginal discs would be more suitable for

use in cell sorting experiments. Each brings its own particular advantages and disadvantages. The main disadvantage to using embryos is that proneural gene expression is very transient when compared to that of imaginal discs. The optimum time period of embryo collection would need to be very highly defined. The SOP cell is present only for a very short moment in time between specification and proneural gene shutdown. In addition specification of SOPs is not synchronised but occurs dynamically over an extended time period which means there is no single ideal time point for collection. Another factor to be considered is that development of SOPs in the embryo is very rapid and therefore GFP expression and activation may lag behind. The advantages of using embryos are that dissociation is technically less complicated than imaginal discs and, given the ease of collecting large numbers of embryos, it should also be possible to obtain sufficiently high numbers of cells without undue difficulty.

The main problem with imaginal disc cells lies in obtaining sufficient quantities of cells, given that the discs must first be dissected from larvae. In addition, harsher dissociation techniques are involved, which may damage the cells and hence change their transcriptional profile. Nevertheless, given the longer time frame when proneural genes are expressed in the imaginal discs, it was thought that the advantages were sufficiently significant to warrant dissociating imaginal disc .

Thus, the initial approach was to isolate specific neural precursor cells from imaginal discs, sort by virtue of the GFP expression and then obtain transcriptional profiles of these cells by using microarray analysis.

4.4 Identification of neural precursor cells

The next question to be considered was how to mark sense organ precursor cells to distinguish them from the surrounding proneural cluster/ectodermal cells. As shown in chapter 2 expression of the proneural genes is under the control of a series of enhancers. Separate enhancers are thought to direct expression independently in the proneural cluster and the SOP cell. A reporter gene, under the control of an SOP specific enhancer, should result in the accumulation of that reporter exclusively in the SOPs. Therefore, it should be possible to utilise the SOP specific enhancers to express green fluorescent protein (GFP) in a subset of neural precursor cells in imaginal discs. The cells of the discs can then be dissociated, and the cells sorted by virtue of the GFP expression. This would then allow a comparison of the transcriptional profiles of different sense organ precursor types or between SOPs and non SOPs for each proneural gene.

4.4.1 Construction of GFP expressing lines

sc-dependent SOPs

There is an enhancer of 356 bp within 3.7 kb 5' from the *sc* gene that promotes Sc protein accumulation in the SOP cells of the wing imaginal disc (Culí and Modolell, 1998).

For specific expression of *sc* in the SOPs of the imaginal discs, I designed primers to the region surrounding the *sc* SOP enhancer and used PCR to amplify this region from genomic DNA. I then cloned this region into pPTGal4 using restriction enzymes BamH1 and EcoR1. This Gal4 vector is a transformation vector that

contains P-element ends, a mini-white selectable marker and a yeast Gal4 under the control of a minimal hsp70 promoter as an enhancer reporter (Sharma *et al.*, 2002).

The *scGal4* construct was microinjected into embryos, transformants were screened for on the basis of eye colour and stable lines were established. I obtained two independent lines for this construct, both on the second chromosome. Flies from each of these lines were then crossed with virgin females possessing a GFP gene under the control of the Gal4 upstream activating sequence (*UAS-GFP*). The pattern of GFP expression correlates with the expression of *sc* (figure 4.4.2). To remove the necessity of carrying out the cross each time, I recombined the *UAS GFP* and the *scGal4*.

ato-dependent SOPs

Sun *et al* 1998 showed that *ato* expression enhancer elements are located 5' and 3' to the *ato* coding sequences (Sun *et al.*, 1998). A series of enhancers located in the 5' region drive tissue specific expression in the chordotonal organ precursors in the embryo and larval leg, wing and antennal imaginal discs. This has been further narrowed down to show that a 367 bp fragment located 3.58 kb upstream of the *ato* gene directed expression in the chordotonal organs of the leg disc only. (zur Lage unpublished observation). The *ato* enhancer was already cloned and injected by P. zur Lage. The line I used in all subsequent experiments had an insertion on the 2nd chromosome. Again I recombined UAS GFP onto the 2nd chromosome.

***amos*-dependent SOPs**

The third proneural gene *amos* does not seem to respond to autoregulation in the same way. In Chapter two I identified a series of enhancers in the *amos* upstream region that appear to control *amos* expression. However the SOP and proneural cluster enhancers do not appear to be as separable as those for *sc* and *ato*. Furthermore, unlike *scGal4, UAS GFP* and *atoGal4, UAS GFP,* expression of GFP in the *amos*-dependent SOPs can be detected only after puparium formation because olfactory neurogenesis is later.

FIG 4.4.1

Utilisation of a GFP reporter gene under the control of specific SOP enhancers resulted in each type of SOP being clearly identifiable by expression of GFP. The next question to be addressed was how best to separate the GFP positive cells from surrounding tissues?

4.5 Cell dissociation of imaginal discs

A number of different methods for dissociating cells from imaginal discs have been published. I carried out experiments to assess these for cell yield and cell viability. All the techniques involve dissection of larvae to remove the discs, followed by enzymatic digestion and mechanical shearing. The cell barriers that resist successful dissociation are the basement membrane and the cell junctions such as septate junctions, adherens junctions and gap junctions which connect the cells together.

I initially dissociated brain/disc complexes, as the dissection of the complexes is easier and less time consuming than dissecting off individual discs. Usually, the brain/discs from 10 larvae were dissociated and the cells counted with a haemocytometer. Dissociated cells were always resuspended in 100µl. In the first method Fehon and Schubiger (1985) remove the cell barriers in a sequential manner. Discs are dissected in Schneider's medium and collagenase is added to unfold the disc and disrupt the peripodial membrane. Citric acid is then added which, as the basement membrane has been previously removed, allows the cells to come apart immediately in the acid. This means the cells need only be exposed to the low pH conditions for 10-15 seconds and so ensures cell survival. Mechanical shearing involved pipetting (using a 200µl eppendorf pipette), 20 times every 10 minutes for the 30 minutes the discs are in collagenase solution, and one occasion of pipetting 12 times following addition of citric acid, and before the mixture is neutralised by addition of Tris base. The pipette tip was coated in BSA before use to prevent any cells sticking to it. A yield of 200 cells per µl was obtained. The cells were resuspended in 100µl of medium, which means a total of 20,000 cells were obtained. Since 10 disc/brain complexes had the potential to provide several million cells 20,000 cells is a very low yield. Therefore, this method not being worth pursuing, was discounted.

The Wyss method (Wyss, 1982) employs protease VIII in ZW medium for enzymatic digestion, followed by pipetting forty times using a 200µl pipette. I determined both the number of cells obtained following dissociation and the percentage of clumping i.e. the number of cells that were not single cells. I carried

out this procedure five times and obtained an average of 1,117 cells / μl following dissociation, with an average of 12.1% of the cells clumping together. Pipetting for an increased number of times had no significant effect on the % of cells that clumped together.

Currie *et al.*, (1988) used dispase and trypsin in Shield and Sang medium, coupled with pipetting 50 times to dissociate the cells. I used this procedure on eight occasions. This method gave an average of 4,674 cells /μl with a similar degree of clumping to that obtained with the Wyss method. As the cells were resuspended in 100μl this means a total of 467,400 cells were successfully dissociated. Of the three methods tested up to this point, the Currie method gave the highest yield (467,400 cells) but, considering the number of cells that could potentially be obtained (several million from 10 disc/brain complexes), the yields from all these methods were very low. It was clear that if cell dissociation were to be a viable method of providing neural precursor cells higher yields would have to be obtained.

A fourth technique involves the use of Trypsin-EDTA (Neufeld *et al.*, 1998). All methods described up to now were developed mainly to study the growth and differentiation of cells in culture. These experiments were also aimed at investigating the ability of different populations of cultured *Drosophila* cells to sort out from each other and the additives needed to promote optimum survival of cells in culture. This technique was the only one that specifically dissociated imaginal disc cells for use in FACS analysis. Neufeld *et al.* (1998) investigated the coordination of growth and cell division in the *Drosophila* wing. In order to characterise the normal

relationship of cell growth to cell cycle progression the authors altered rates of division in cell clones or compartments of the *Drosophila* wing. They then measured the effects on growth by dissociating wing discs into single cells and using an analytical cell sorter to determine cell numbers, DNA content and cell size. This technique was used subsequently to show that Ras1 promotes cellular growth in the *Drosophila* wing (Prober and Edgar, 2000) and to study the interactions between Ras1, dMyc and dP13K signalling in the wing (Prober and Edgar, 2002).

Brain/discs complexes were dissociated in 450µl Trypsin-EDTA and 50µl 1X PBS for two hours. Following enzymatic digestion, the sample was pipetted 40 times. In my hands this technique proved the most successful as far as cell yield was concerned, giving an average of 16,082 cells / µl (1,608,200 in total) from seven experiments. This was a marked improvement on the yields obtained by other methods and as such became the method of choice. Visual inspection showed that cells were rarely found in clumps. Compared to the Currie method, which gave the second highest yield there was a four fold increase in the number of cells obtained when using Neufeld *et al.* (1998). It is nevertheless clear that large numbers of cells are lost in the procedure, as only 1.6 million dissociated cells are obtained from a starting population of several million. This is borne out by figures obtained from dissociating discs only.

When only discs rather than disc brain complexes were dissected, using the Neufeld method, the leg and wing discs from ten larvae gave an average of 4,309 cells /µl or 430,900 in total from seven experiments. Wing discs contain 50,000 cells

and leg discs 20,000 cells. Ten larvae contain a total of 20 wing discs and 60 leg discs. However 80 discs were never successfully dissected. I would estimate that there were perhaps 15 wing discs and 40 leg discs present per sample. This means there is a total of approximately 1.5 million cells from which only 430,900 were successfully dissociated. The dissociated cell population contains both *sc*-dependent neural precursors and *ato* dependent precursors.

As there is a large cluster of *ato* expressing cells in the disc for the femoral chordotonal organ (FCO) I next dissociated leg discs only. This resulted in an average yield of 1,718 cells/μl from three experiments. Given that cells were always resuspended in 100μl of medium this meant 171,800 cells on average were obtained from dissociating leg discs from ten larvae. Ten larvae contain a total of 1.2 million leg disc cells. Even allowing for the fact that six discs were probably never successfully dissected from each larva this means that a large number of cells are lost in the procedure. If for example, four discs are dissected from each larvae this means three quarters are lost in the dissociation process. Although this is still a high proportion of cells to lose, the cell dissociation technique of Neufeld *et al* 1998, proved by far the most successful.

FIG 4.5.1

Wyss discs/brains	Fehon discs/brains	Currie discs/brains	Neufeld discs/brains	Neufeld discs Leg/wing	Neufeld discs leg
590	200	2,260	20,000	2,470	1,020
230		6,125	8,810	8,075	2,133
255		6,040	6,760	2,587	2,000
1,845		4,560	12,000	4,167	
2,665		5,395	10,000	4,100	
		7,350	40,000	4,767	
		2,660	15,000	4,000	
		3,000			
1,117		4,673.8	16,081.4	4,309.4	1,717.7

Table 4.5.1 Comparison of the number of cells obtained per µl using the different methods of cell dissociation.
The first four columns refer to dissociation of disc/brain complexes, column five refers to dissociation of wing and leg discs, and column six to dissociation of leg discs. All figures relates to 10 larvae being dissected. The last row is the average of cells obtained with each dissociation method.

4.6 RNA extraction and RT- PCR

One concern was that the harsh enzymatic conditions to which the dissociated cells were exposed might interfere with RNA quality. Therefore, I extracted RNA from dissociated cells to use in RT-PCR experiments.

4.6.1 RNA extraction using "Cells to cDNA" followed by RT-PCR

Initially an Ambion product called "Cells to cDNA" was used to extract RNA. This procedure allows reverse transcription (RT) without RNA isolation. Cells are subjected to a heat treatment in Cell Lysis Buffer which lyses the cell membranes and releases the RNA into the solution. Simultaneously, the heat step inactivates the endogenous RNases. Following DNase treatment, reverse transcriptase is added and PCR performed. Although this method is suitable for small sample numbers, and is quicker than standard methods that involve first isolating RNA before reverse transcription, this technique never worked satisfactorily. Contamination from genomic DNA was frequently present in samples analysed. I used primers to a number of different genes, including *GFP, sc, ato, amos, sens* and *ed.* The reverse transcriptase used was from M-MLV. Control PCR reactions did not contain the reverse transcriptase and so a product should not be present unless there is genomic DNA contamination. Although at times I observed a band of the expected size in the lane loaded with the reaction containing the reverse transcriptase, and absent in the corresponding lane without the reverse transcriptase, reproducibility was extremely poor.

4.6.2 RNA extraction using Trizol followed by RT- PCR

Due to the unreliability of the "Cells to cDNA" system I decided to employ an alternative method of RNA extraction. Treatment of the RNA with DNase I successfully eliminates DNA contamination in this case. cDNA can then be made

using a reverse transcriptase which can then be used in PCR reactions. As detailed below, this proved to be a much more reliable system.

I extracted RNA from intact and dissociated imaginal discs using Trizol and carried out reverse transcriptase reactions to make cDNA. The reverse transcriptase enzyme I used was Superscript from Invitrogen. Primer pairs to *GFP*, *ed* and *ato* were used to assess the quality of RNA obtained and initial contamination problems were overcome.

RT-PCR using RNA extracted from intact leg discs

The first issue to be addressed in this context was, is it possible to detect small amounts of RNA. RNA was extracted from intact leg discs from 15 *atoGal4, UAS GFP* larvae and reverse transcription reactions using primer pairs to *ed,* GFP and *ato* were carried out. Presuming all 6 leg discs were dissected off, the maximum number of cells present could have been 1.8 million. In reality, there may have been half this number of discs so that 4,800 could have been chordotonal precursor cells (80 scolopidia per leg disc) and hence *ato* and GFP expressing. The RT-PCR reactions detected a band of the expected size for all three primer pairs. Contamination was absent from the control reactions. The ability to detect GFP and *ato* in a population of 4,800 SOP cells suggests this is a very sensitive technique.

FIG 4.6.1

RT-PCR using RNA extracted from dissociated leg discs

The second question to be addressed concerns the quality of RNA extracted from dissociated cells, i.e. is the quality comparable to that extracted from intact imaginal discs or does the Trypsin-EDTA used to dissociate the cells damage the

RNA in some way. Again, leg discs were removed from 15 larvae of the *atoGal4, UAS GFP* line but in this case enzymatic digestion and dissociation was carried out. The cells obtained were counted and the total number was 625,000. If, for example, four discs from each larvae are present in the sample, this is a yield of 52%. This is an exceptionally high yield for dissociation experiments. This suggests approximately 2,400 were neural precursor cells. RNA was extracted and RT-PCR again carried out using primer pairs to *ed*, *ato* and *GFP*.

FIG. 4.6.2

There appears to be no discernable difference between RNA extracted from dissociated cells and RNA extracted from intact imaginal discs and therefore, it can be concluded that the dissociation process does not damage the quality of RNA subsequently extracted from these cells.

The aim of this section was to find a reliable method for dissociating imaginal discs into single cells. The technique involving trypsin-EDTA as described by Neufeld *et al.* (1998) gave the highest yield and as such became the method of choice. RT-PCR experiments were carried out to ensure the dissociation process did not compromise the quality of RNA extracted.

Dissociation results in a mixed population of neural/non-neural cells. Consequently, the population must now be subdivided into neural/non-neural cells. As the neural precursor cells express GFP under the control of the SOP enhancer, this can be accomplished by utilising fluorescent activated cell sorting (FACS) to sort the cells by virtue of GFP expression.

4.7 Cell sorting

The technique of flow cytometry is extremely effective, both when used solely in an analytical manner (for example when sorting mammalian leukocyte subtypes) and more effective still when employed physically to separate different subpopulations of cells by using GFP expression.

Fluorescently labelled mammalian cells are commonly analysed by flow cytometry. Malatesta *et al.* (2000) fluorescently labelled radial glial cells and isolated them via FACS. Radial glial cells are found in the central nervous system of vertebrates and the authors demonstrated that isolated cells generate both neurons and astrocytes. It proved possible to culture these cells illustrating the ability of cells to survive the sorting process (Malatesta *et al.*, 2000) Aubert *et al.* (2003) used a reporter mouse line in which GFP was inserted into the *Sox1* locus. *Sox1* is a transcription factor found in mammalian neural progenitors. Embryonic day 10.5 Sox1-GFP positive embryos were digested in trypsin and the cells sorted by FACs to give Sox1-GFP positive and Sox1-GFP negative cells. RT-PCR to known genes confirmed the efficient separation of neural and nonneural cell populations. Several differentially expressed genes were identified by hybridisation to a custom array, thus illustrating the suitability of sorted cells for use in microarray experiments.

4.7.1 Cell sorting in *Drosophila*

There is a number of precedents for FACS of *Drosophila* cells using GFP to mark the cell of interest. Bryant *et al.* (1999) used this approach successfully to isolate somatic follicle cells from *Drosophila* ovaries with a view to identifying genes expressed in the follicle cells with potential roles in axis formation. Spatially restricted subpopulations were marked for FACS selection using a GFP reporter and follicle cells were then purified from expressing and non-GFP expressing, and inspection of the sorted cells showed a greater than 90% purity of the visibly GFP positive cells. To verify that the cells had not been damaged by enzymatic digestion, slot blots were carried out for the reporter construct (GFP) and for an endognenous posterior cell-specific transcript (*pointed* P1). These transcripts were enriched as expected in the cDNA amplified from the posterior GFP expressing follicle cells. This demonstrated satisfactorily that cells isolated by cell sorting maintain their original fates. Thus, in principle, the techniques can be readily applied to imaginal disc and hence to SOPs.

4.7.2 Cell sorting using GFP expressing lines

I used a Cytomation Moflo cell sorter installed in the Institute for Stem Cell Research. This machine is specifically geared to the detection of GFP expressing cells. However up to this point in time it had never been used to sort *Drosophila* cells. *Drosophila* cells are smaller than mammalian cells and I thought some optimisation might be necessary in the beginning and so I started experiments using a line that expresses GFP widely. The results of cell sorting experiments will be presented as scatter plots.

Initial experiments focused on determining the percentage of GFP positive cells in each of the GFP expressing fly lines. I began cell sorting experiments using a line (*lzGal4, UAS GFP),* that expresses GFP in the brain as well as in eye-antennal discs at the third instar larval stage. Judging from the expression pattern this line should possess considerably more GFP positive cells than either *scGal4, UAS GFP* or *atoGal4, UAS GFP*. My first cell sorting experiment involved dissecting brain/discs complexes from ten *lzGal4, UAS GFP* larvae and dissociating them according to the method of Neufeld *et al.* (1998). Following cell dissociation, the cells were filtered through a 40 μm mesh in preparation for cell sorting. The sample was transported on ice to the Institute for Stem Cell Research and immediately put through the flow cytometer and sorted into GFP positive and GFP negative cell populations. There were approximately 1 million cells before sorting and 3.7% of these were GFP positive according to the cell sorter. However, of the starting population a proportion will die and others remain in the cell sorter, so the entire starting population can never be recovered. I repeated the *lzgal4, UAS GFP* experiment on two further occasions. The percentage of GFP positive cells was 3.03% and 2.27%. Therefore, from three experiments carried out using *lzgal4, UAS GFP,* on average 3% of cells were GFP positive.

The initial experiments using *lzGal4, UAS GFP* demonstrated successfully that the cell sorter was suitable to use in sorting *Drosophila* cells. I then proceeded to determine the number of GFP positive cells, according to the cell sorter, in both *scGal4, UAS GFP* and *atoGal4, UAS GFP*. GFP is expressed in the *sc*-dependent SOPs of the wing disc for the *scGal4, UAS GFP* line. For the *atoGAL4, UAS GFP*

line, GFP is expressed in the *ato*-dependent SOPs of the femoral chordotonal organ (FCO) and perhaps in the tibial chordotonal organ (TCO) of the leg disc. I carried out a cell sorting experiment with *scGal4, UAS GFP* on one occasion and *atoGal4, UAS GFP* on five separate occasions. I concentrated on the *ato* specific SOPs in the leg disc for the following reasons:

1. The microarray experiments I carried out on whole embryos had focused on identifying downstream targets of *ato*. If cell sorting proved successful putative downstream targets identified could be compared with those found by microarray analysis on whole embryos.
2. The *atoGal4, UAS GFP* is expressed in a large cluster of some 70-80 SOPs in the leg disc that give rise to femoral chordotonal organs. This large cluster is formed by the interplay of EGFR and Notch signalling (zur Lage and Jarman, 1999). The *scGal4, UAS GFP* is expressed in the *sc*-dependent SOPs of the wing disc. There are approximately 20 proneural clusters in the wing disc (Gomez-Skarmeta *et al.*, 1995) and from each of these two SOPs are selected. This means that in each wing disc of genotype *scGAL4, UAS GFP*, there are approximately 40 GFP expressing cells. Given the greater number of GFP positive cells in the leg disc under the control of *ato*, it appeared sensible to concentrate on the *atoGal4, UAS GFP* line.

This was borne out by the number of GFP positive cells detected by the cell sorter. For *atoGal4, UAS GFP* the percentage of GFP positive cells ranged from 0.22% to 1.79%. The average over five sorts was 0.86%. The wide range of GFP positive cells can be explained by the fact that formation of SOPs in the FCO takes place over time as those formed first signal back to recruit more. Therefore leg discs

from slightly older larvae will contain a higher number of SOPs and therefore a higher number of GFP positive cells. In addition, some SOPs may have divided resulting in a higher number of GFP positive cells in some instances. I also dissociated and sorted wing disc cells from the *scGal4, UAS GFP* line. As expected the *scGal4,UAS GFP* line contains fewer GFP positive cells than the *atoGal4,UAS GFP* line. The cell sorter gave a reading of 0.41% GFP positive cells.

The cell sorter gives a higher value for the percentage of GFP positive cells than might be expected. For example, for the *atoGal4, UAS GFP,* one would expect only 0.004% of cells to be GFP positive assuming 80 scolopidia per leg disc (this number may vary slightly depending on age). However the cell sorter gives an average count of 0.86%. This is a considerable difference in yield and suggests that either there are more chordotonal precursors in the leg disc than thought or perhaps that SOPs preferentially survive cell dissociation when compared to the other cells of the leg disc.

LzGal4, UAS GFP	*scGal4,UAS GFP*	*atoGal4, UAS GFP*
3.03	0.41	0.35
3.7		0.70
2.27		0.22
		1.28
		1.79

Table 4.7.1 Percentage of GFP positive cells found in each of the fly lines.

4.7.3 Cell sorting control experiments

As with any cell sorting experiments, it is necessary to choose the cells of interest i.e. to gate upon the required cell population. We (the cell sorter technician and I) naturally chose the cells based on fluorescence i.e. those with the highest level of fluorescence were taken to be the GFP positive cells. I decided to confirm this by dissociating discs and sorting the cells from a non-GFP expressing line in tandem with a GFP expressing one. Therefore I dissociated and sorted cells from *Oregan R* imaginal discs and compared the profile obtained to that obtained from sorting cells from *atoGal4, UAS GFP*. The GFP positive cells from the *atoGal4, UAS GFP* line are present in a region to the lower right of the profile. This region is almost completely empty of cells in the OrR profile indicating that only GFP positive cells were being gated upon.

FIG. 4.7.1

To determine the number of living cells, I carried out some experiments using propidium iodide, a fluorescent dye which is taken up by dead cells. The profile obtained showed that most of the cells were alive on entering the cell sorter (figure 4.7.2 A). Those in region R4 were dead, i.e. only 1.31% of the total cell count. Presenting the data from a subsequent experiment in a slightly different way showed 89.32% of the cells to be in region R3 (figure 4.7.2 B). These cells are alive and GFP negative. Region R5 contains the GFP positive cells and these are 1.21% of the total. Cells between these two regions are alive but cannot be unambiguously assigned to either region on the basis of GFP expression and cells above these two regions have died.

FIG 4.7.2

Cell sorting results in two separate populations of cells, those that are GFP positive and those that are not. The GFP positive cells are neural precursor cells. Control experiments involving a non-GFP expressing line, leaves me confident that the correct population of cells is being gated upon and propidium iodide staining suggests the majority of the cells are alive upon entering the cell sorter.

4.7.4 Limitations encountered

A critical factor in cell sorting is the time required to dissect and remove the discs from a sufficient number of larvae. Clearly this must be done as quickly and efficiently as possible to reduce the time taken to prepare the cells for sorting. In order to overcome this, fellow lab members dissected larvae simultaneously and in one hour 450 leg discs were obtained. These discs were dissociated and the total cell count on a haemocytometer was 1.4 million cells and, as 450 leg discs contain 9 million cells in total, the percent yield was 15.5%. This was a lower yield than expected as usually about 75% is lost in the dissociation procedure.

The cell sorter gave a count of approximately 900,000 cells and when sorted 1.28% of these cells were positive. Theoretically this would mean 11,500 cells were GFP expressing neural precursor cells and it should easily be possible to extract RNA from this number of cells. RNA extraction experiments I carried out on dissociated but not sorted cells gave, on average, 5.6μg of RNA from 640,000 cells. This suggests in my hands, 0.1μg of RNA would be obtained from the 11,500 sorted cells. 12μg of RNA (with a lower limit of 5μg) is usually required as the starting material for microarray analysis unless a second round of amplification is used. At

the time of these experiments, such amplification procedures had not been well established. A possible way around this problem would be to store cells in Trizol and pool material from different sorts before extracting RNA.

There is clearly a difficulty in obtaining sufficient RNA for microarray analysis and this will be returned to in the next paragraph. In any case, I needed to determine if the process of cell sorting was deleterious to the cells. Therefore, I extracted RNA from 60,000 sorted GFP negative cells. In tandem, I extracted RNA from 60,000 dissociated but not sorted cells and performed RT-PCR on both samples using *histone2A* primers. I chose this gene to start with because it is very highly expressed. Both samples gave a band of the expected size suggesting that cells maintain their integrity following sorting (although the band present in the sorted cells lane is fainter). I then followed up this experiment by carrying out RT-PCR on the sorted cells using primers to *ed* and to *sens*. I obtained a faint band of the expected size for *ed,* and nothing for *sens*. It is not surprising that no band is observed for *sens,* as the cells are GFP negative. This means they are not SOP cells and *sens* is expressed exclusively in SOPs. The RT-PCR experiments suggest that, following cell sorting, the cells in principle are suitable for use in downstream application such as microarray analysis.

FIG. 4.7.3

Although the principle behind sorting GFP positive cells from imaginal discs appears to be a sound one, I had doubts about the practicalities of the approach. More specifically, I was concerned about the need to dissect larvae to remove the discs. I felt that this was the rate-limiting step and that it may not be possible

because of this, to obtain sufficient cells to make sufficient RNA for microarray analysis. From experiments I carried out on dissociated but not sorted cells, I estimate 570,000 GFP expressing SOP cells would be necessary to obtain the 5μg of RNA required for microarray analysis. If 0.1μg of RNA is obtained from 11,500 cells, this indicates that the mass dissection procedure would have to be carried out 50 times to obtain sufficient GFP positive cells. Although in theory possible, as cells could be stored in Trizol until needed, this seems a somewhat daunting process. There is scope for improvement in RNA extraction procedures, on one occasion for example, I obtained 18.24μg from 388,000 cells. Extrapolating from this figure suggests that 11,500 cells would give 0.5μg of RNA. In that case, the mass dissection would only need to be repeated 10 times rather than 50. However, there is no guarantee that I would be able improve my RNA extraction technique to the extent that it would routinely yield this amount. Therefore, given the difficulties inherent in obtaining the large numbers of SOPs required to make RNA, I decided to explore the possibilities of dissociating embryos, the advantages being that dissection of embryos is not required and that it is possible to start with a much greater amount of material.

4.8 Embryos

4.8.1 Obtaining a GFP expressing line

An *atoGal 4* line expressing Gal4 in the embryonic *ato*-dependent SOPs cells was obtained from Bassam Hassan (University of Leuven, Belgium). GFP expression is observed in all *ato*-dependent cells including chordotonal, antenno-

maxillary complex and Bolwig's Organ. The line consists of a 3.6 kb fragment directly upstream of *ato* in a GAL4 vector. (Hassan *et al.*, 2000). The *atoGal4* line was recombined with the *UASGFP* to remove the necessity of having to carry out the cross each time.

4.8.2 Determining optimum time of GFP expression

The optimum time of GFP expression was determined by carrying out a time course. I initially allowed the flies to lay at 25 °C for one hour, and aged the embryos for a further 6.5 hours. This time schedule proved incompatible with the hours of operation of the cell sorter and I therefore altered my methods of collection. Flies in cages were allowed to lay at 18° for two hours, embryos collected and aged for 13 hours. This corresponds approximately to embryos being at a developmental age of between 6.5–7.5 hours 25°C.

ato expression initially occurs in proneural clusters before being refined to single cells. Single cell expression is first observed in the P cell (most dorsal of the *ato* expressing SOPs) and is lost from this cell before expression is switched on in the remaining *ato*- dependent SOPs. By 6.5-7.5 hours *ato* expression is refined to four single cells in each segment. This means that at this time, *ato* expression has already been lost from the P cell. There are delays due to Gal4 and to GFP. Therefore at 6.5-7.5 hours GFP is expressed in the P cell and in either one or two of the other *ato*-dependent SOPs. GFP expression will follow later in the remaining *ato* expressing cells. As GFP is not turned on in all *ato* expressing cells at the same time, a compromise time must be arrived at whereby the greatest number of cells

possible are expressing GFP and *ato* simultaneously. As GFP perdures after *ato* expression has been switched off, it is important also not to choose too late a time point. 6.5-7.5 hours appeared suitable to use as a time point. Further time points should be carried out to confirm this, and the possibility of using a tighter time frame for collection should also be investigated.

A complication encountered was that the embryonic enhancer was thought to be an SOP specific one but this proved not to be the case. Sun *et al.* (1998) studied the embryonic expression pattern of this enhancer by cloning it into a transformation vector containing *lacZ* as a reporter gene. The authors state that fusion of either 2.6 kb or 5.1kb of ato 5' sequences to *lacZ* results in *lacZ* expression exclusively in the SOP cells of the chordotonal organs. The fragment I used is 3.6 kb in length and I had expected it to display a similar expression pattern. However, my experiments clearly show GFP in proneural cluster cells between 6.5 and 7.5 hours after egg laying. This suggests there may either be both a SOP and a proneural cluster enhancer element present in this 5' region or perhaps a single enhancer that is capable of driving GFP expression in both SOPs and PNCs. GFP perdurance can be used to follow the fate of these cells and staining of older embryos with 22C10 (Hummel *et al.*, 2000) shows GFP present in ectodermal cells and also overlapping with chordotonal organs. This strongly indicates the presence of a proneural cluster enhancer.

However, given the rapid and transient nature of formation of SOPs in the embryo, the fact that this enhancer is a proneural cluster one as well as an SOP one may be an advantage. The SOP stage is very transient in the embryo and it may not

be possible to pinpoint the precise moment in time when this happens. A true SOP-specific enhancer would probably not become visible until late in the life of the SOP, and overlap little with proneural gene expression. This is because of the delay between the synthesis of GFP and it becoming detectable fluorescently following cyclisation. This delay is several hours, (Kumar *et al.*, 2003), which is a long time in the life of an embryo. The presence of GFP in the PNC gives a wider window of opportunity to sort the correct cells as obviously there are SOP cells within each proneural cluster.

FIG. 4.8.1

4.9 Embryo dissociation

Embryo dissociation involves dechorionating the embryos followed by homogenizing in a Dounce homogenizer. Unlike dissociation of imaginal disc cells no enzymatic treatment is required (Allis *et al.*, 1977). From a staged two hour collection (one cage containing approximately 500 flies) following dissociation 550,000 cells were obtained. RNA was extracted from these dissociated cells and 2 μg of RNA obtained. The problem of limiting availability of imaginal disc cells was not an issue as far as embryonic cells were concerned because large numbers of embryos can be dissociated in a short time. In fact, the only limiting factor was the number of cages which could be coped with at once.

FIG. 4.9.1

4.10 Embryonic cell sorting

As mentioned previously, I recombined the *atoGal4* line with *UAS GFP*. While awaiting completion of these crosses, I performed FACS on embryos of phenotype *atoGAL4 x UAS GFP*. Embryos were collected for 1hr at 25° C and then aged for a further 6 hr 30 minutes. In parallel, I collected and aged embryos from *OrR* flies. I did this because I had two main aims for this first embryonic cell sort

(i) to determine the typical profile for GFP positive embryonic cells and

(ii) to obtain enough sorted cells to successfully isolate RNA.

I decided to collect, dissociate and sort *OrR* embryos as well.

From the *atoGal4 X UAS GFP,* 1.23% of the cells were GFP positive. However to my surprise, 0.64% of the OrR cells registered as GFP positive. Clearly, this is an impossibility and this result suggested that some cells in the *atoGal4 X UAS GFP* cell population, which the cell sorter was registering as GFP positive may in fact be GFP negative.

The fact that 0.64% of the *OrR* cells registered as GFP positive gave cause for concern. To ensure that the correct region cell population was being selected, i.e. being gated upon I carried out a number of control sorts. I sorted dissociated *OrR* cells and also dissociated cells from a line of flies that express GFP in all cells of the embryo. The region (gate) previously selected as the one where GFP positive cells were present contained 67.45% of the cells from the line that expresses GFP ubiquitously. The corresponding region on the *OrR* plot should have been empty (no GFP positive cells are present in *OrR* embryos), but 0.52% of cells were found here, suggesting that the parameters used to select the GFP positive cells should be moved slightly.

FIG. 4.10.1

I carried out one final cell sort on dissociated embryonic cells bearing in mind the results from the control experiments and hence altering the "gate" parameters slightly. I used the recombined line *atoGal4, UAS GFP*. For the *atoGal4 X UAS GFP* sort, I used embryos from a one hour collection at 25° C and aged for a further six hours 30 minutes. However the embryos used for this sort had been laid over two hours at 18° C and aged for a further 13 hours. On this occasion, 1.5% of the cells registered as being GFP positive. This is a higher proportion of GFP positive cells than that obtained previously (1.23%), even when using more stringent selection criteria. This can be explained by the different methods used to age the embryos. A purity check was carried out to ensure the cells were really GFP positive or GFP negative. This involves running a proportion of the sorted cells through the machine a second time. For the GFP positive cells, 92.59% of the cells were present in the region previously selected as the GFP positive region, whereas 6.67% were not. For the previously selected GFP negative cells, 94.77% were in the correct region to be GFP negative (the remainder being just outside the boundary) and no cells were found in the GFP positive region of the scatter plot. I could be confident that the cells had been separated successfully into GFP positive and GFP negative cells. Future cell sorting experiments involving embryonic cells, should, therefore concentrate on the numbers of cells obtained from each embryonic collection and on optimising the cell sorting procedure to ensure the greatest number of GFP cells possible are obtained.

FIG 4.10.2

Cell sorting was used successfully to separate GFP positive and GFP negative cells. For imaginal discs, the enhancers used to drive GFP are expressed specifically in the neural precursor cells. This means all GFP positive cells are neural precursor cells. In the case of the embryo, the enhancer used expresses GFP in the SOPs and also in some of the proneural cluster cells. Therefore the embryonic GFP positive population consists of both SOPs and PNCs. This may in fact be an advantage, as explained in section 4.8.3.

FIG 4.10.3

4.11 RNA Extraction

From my first sort I obtained $2x10^6$ cells from the OrR line. I extracted RNA from these cells and obtained 1μg of RNA instead of the expected 15 μg. Nevertheless, I carried out RT-PCR using *histone2A* primers and obtained a band of the expected size. I also extracted RNA from the 85,000 GFP negative cells sorted from *atoGal4 X UAS GFP*, and obtained 0.3μg (expected yield approximately 0.7μg). I again used *histone* primers in an RT-PCR reaction but could not detect a band. I also used *ato* primers but the primers appear to be contaminated.

4.12 Discussion

This part of my project arose from my doubt that whole embryo microarray experiments would be able to provide all the answers in identifying proneural target genes. The problems inherent in using whole embryos are clearly illustrated when one considers the percentage of GFP positive cells, i.e. if only 1.5% of the cells in a whole embryo are neural precursors it may not be possible to detect transcripts of

some of the more weakly expressed genes. Therefore, my aim was to explore cell sorting as a way of comparing the transcriptional profiles of different sense organ precursor cell types. The immediately obvious advantage to dissociating either imaginal discs or embryos and sorting cells by virtue of GFP expression when compared to whole embryo analysis, is that a pure population of sense organ precursor cells is obtained. This means that the chances of discovering differentially expressed genes in these cells when compared to non-neural precursors are greatly increased.

The limiting factor concerning the disc cells was one of quantity given that larvae had to be dissected to remove the imaginal discs. I believe, however, that it would be possible to obtain sufficient RNA over time if the sorted cells were stored in Trizol at -80° C. Given the percentage of GFP positive cells and the numbers needed to make RNA, this is not something to be under-taken lightly. Six people dissecting for one hour produced a total of 450 leg discs. As each leg disc contains 20,000 cells this should mean there are potentially 9 million cells to be dissociated. Following dissociation the cell count as estimated from a haemocytometer was 1.4 million cells. The FACS machine estimated 900,000 of which 1.28% i.e. 11,500 were GFP positive. While, as shown by the propidium iodide staining, most of the cells are alive upon entering the cell sorter, all will probably not be recovered because some will become stuck to the wall of the cell sorter and others to tubes and pipette tips. Even if we assume, for now, that 11,500 viable cells will be recovered, this still means huge numbers of cells are lost in the procedure. 11,500 cells recovered from a potential starting quantity of 9 million means only 0.12% of cells are the cells of interest.

Extrapolating to one person dissecting for one hour (it not being considered wise to dissect for longer than this with the aim of maintaining the same transcriptional profile for all discs), indicates the likelihood that 75 leg discs would be dissected. This implies that only about 2,000 GFP positive cells would be obtained from a single sort. Clearly, this is an insufficient quantity for my purpose, bearing in mind, that ideally 12 μg of RNA are required for microarray analysis. Although it is possible to make do with less, 5 μg being just about acceptable for the standard procedure, anything less than that would require a second round of amplification. As mentioned previously, I estimate that 570,000 cells should be sufficient to obtain 5 μg of RNA at my current rate of RNA extraction. To obtain 570,000 SOPs, 21,375 leg discs would need to be dissociated (450 leg discs = 11,500 SOPs). Even with the assistance of fellow lab members, it would be difficult using the methods outlined above, to obtain enough RNA to make the micorarray worth attempting.

4.12.1 Ways to improve the viability of dissociating imaginal discs

It had become clear that if the procedure of dissociating and sorting imaginal disc cells were to be successful, the above approach would have to be amended. Two possibilities come to mind:

a) The dissection process at the very beginning of the procedure, whereby larvae are dissected and the discs removed, limits the amount of available starting material. One big improvement to the procedure would be if more discs were available to dissociate. For instance by scaling up the isolation of imaginal discs through the use

of a density gradient to separate them from surrounding tissue. This would involve collection of larvae from a population cage followed by grinding the larvae suspension to release imaginal discs from attachments to the larval epidermis. Filtration to remove larval carcasses, gravity settling to separate dense (i.e. imaginal discs) from less dense (i.e. fat) tissues, followed by a Ficoll step gradient and differential adhesion of disc tissue to glass surfaces would allow mass isolation of discs. This method has not been explored further but should not involve undue difficulty.

b) Another possibility would be to carry out a second round of amplification on the RNA extracted from the cells. It has been shown that a second round of amplification can be used to obtain sufficient RNA from small sample sizes such as biopsies, laser captured microdissection of tissues (Ernst *et al.*, 2002; Ohyama *et al.*, 2000) or flow-sorted cells (Posakony UCSD, Unpublished, abstract presented at CSHL, Symposium on neurobiology of *Drosophila*, 2003). This involves carrying out two cycles of standard cDNA synthesis and *in vitro* transcription (IVT). The first cycle provides initial amplification of total RNA, resulting in unlabeled cRNA. In the second cycle of IVT synthesis, biotin-ribonucleotides are incorporated to produce labelled antisense cRNA target. This can be carried out successfully using 100 ng to less than 10 ng of RNA. The main concern with this procedure is that only certain transcripts may be amplified; however, work carried out by Affymetrix research and development department suggests that this is not the case. Transcripts amplified by the small sample procedure were compared with transcripts amplified by the standard assay. There appears to be no significant difference in transcripts labelled by the two methods.

Not only is it possible in principle to obtain sufficient RNA from extremely small samples but personnel at the Sir Henry Wellcome Functional Genomics Facility have begun using this procedure in practice. They now routinely amplify 50-100ng of total RNA using the small sample procedure. In addition, they have successfully amplified sufficient RNA from 0.08 cubic mm of cells although they were unable to detect a total RNA reading on the spectrophotometer or on the Agilent Bioanalyzer. Given the relatively recent advances in technology and the practical experience held by the Functional Genomics Centre, the small sample protocol could be used to obtain sufficient RNA from very few GFP positive cells. Extrapolating from previous experiments suggests that 11,400 cells would be sufficient to obtain 100 ng of RNA, which should be more than adequate to carry out microarray analysis using the small sample procedure. As 11,500 GFP positive cells were previously obtained from 450 leg discs this is most certainly achievable.

Indeed members the Posakony lab at UCSD recently presented an abstract at CSHL, Symposium on neurobiology of *Drosophila,* 2003, where they outlined the progress they have made in profiling the gene expression of PNCs in the wing disc. They utilised a GFP reporter gene under the control *enhancer of split m4*. This gave GFP expression specifically in the PNCs of the wing disc. Wing discs were dissociated and cells sorted by virtue of GFP expression. Each wing disc contained an average of 1,000 GFP expressing cells. This made it easier for them to obtain sufficient numbers to extract RNA from by using the small sample procedure. Microarray analysis was carried out in triplicate on PNC cells versus non-PNC cells. Of the candidate genes identified, 24 were previously known to be expressed in PNCs and/or SOPs. Expression analysis was carried out on 41 new candidate genes,

and of these, 9 are expressed in PNC cells and 15 specifically in SOPs. This demonstrates that the principle behind the approach outlined in this chapter is sound.

Other avenues used when studying gene expression profiles in imaginal discs have involved microdissection of discs to isolate the area of interest. Butler *et al.* (2003) cut between the presumptive hinge and body wall regions to give two complementary wing disc fragments. The authors then used microarrays to identify genes that are differentially expressed in the proximal (body wall) and distal (wing blade) regions of the disc. As these regions are separated by a lineage restriction that occurs in first larval instar stage and are relatively large, this was an ideal approach. Francoise Chanut (UCSF), presented an abstract at the CSHL 2003 meeting, detailing her microdissection of eye discs into pre-furrow, morphogenetic furrow and post-furrow regions with the aim of determining which genes are expressed specifically in the furrow. That being said, it is not really practical in the case of SOPs except possibly for *ato* expression in the leg disc when a large cluster of GFP positive cells is formed by Egfr signalling (zur Lage and Jarman, 1999). It may be possible to remove this cluster from surrounding tissue by microdissection thus removing the need for cell sorting. Of course all the microdissection techniques require two rounds of RNA amplification.

4.12.2 Embryonic Cells

Dissociation of embryonic cells appeared to be eminently feasible as a means of obtaining pure populations of neural precursor cells which could then be used in microarray analysis to determine differentially expressed genes. The main advantage

here was that dissection was no longer necessary. On the other hand, there was the difficulty that the embryos had to be collected within a very short time-frame to ensure that they were all approximately the same age. An *atoGAL4, UAS GFP* line was used initially to examine the feasibility of this approach. Optimum age was determined and cells were successfully dissociated and sorted. RNA was made from sorted cells on one occasion but a very low yield was obtained. Due to time constraints this was not repeated but estimates were made using dissociated cells as to how many cells would be needed to obtain sufficient RNA for microarray analysis.

As I have described, flies were allowed to lay in a cage at 18° C for two hours. The embryos were dissociated and counted on a haemocytometer. 500,000 cells were obtained from this collection. RNA was extracted from these cells and 2 μg was obtained. This is lower than that obtained when extracting RNA from dissociated imaginal disc cells (500,000 cells yields 5 μg on average), and it should therefore, be possible to improve the yield of RNA. However the time frame did not allow this and as a consequence, the following calculations are based on obtaining 2 μg of RNA from 500,000 embryonic cells.

Microarray analysis if carried out using the standard procedure requires a minimum of 5 μg of RNA, entailing starting with 1.25 million GFP positive cells. The percentage of GFP positive cells in *atoGal4, UAS GFP* embryos, when laid at 18 °C and aged for between 13 and 15 hours is 1.5% (based, however only on one experiment). It follows that, to obtain 1.25 million GFP positive cells it would be

necessary to start with 84 million dissociated cells. A single "two hour" collection resulted in 500,000 cells; this implies 166 individual collections are necessary to obtain the required number. Obviously not all collections will result in the same amount of embryos but, taken as a whole, this procedure should be sufficiently productive. It is clearly impossible to collect and dissociate the contents of 166 cages within the required time. To overcome this problem, it would be necessary to collect and dissociate the embryos in successive batches, and, where appropriate, store the embryo batches in Trizol. It should be remembered that frequently large numbers of cells are lost in the sorting procedure and more collections may be necessary to obtain the required number. I feel that not more than 10 cages should be collected from at the any one time and as such this would mean 17 different sessions. This would mean three and a half weeks of one session a day. This could of course be reduced if more people were pressed into service for embryo collections/dissociations as the flow cytometer is capable of sorting large numbers of cells. In any case this is a substantial undertaking especially as each microarray experiment should be repeated at least four times to ensure reproducibility of results.

Although obtaining sufficient GFP positive embryonic cells should be easier than obtaining sufficient imaginal disc cells, clearly it will still require a large input of time and effort to obtain sufficient RNA to carry out microarray experiments by the standard procedure. Alternatively, given Glasgow's recent success with the small sample procedure, it would of course be possible to use the small sample procedure to extract RNA from dissociated embryonic cells. If 500,000 cells yielding 2 μg of RNA is taken as a base line, then 100 ng should be obtained from 25,000 cells and, 25,000 GFP positive cells should be recovered from a starting population of 1.5

million cells. As one two hour collection of embryos provides 500,000 cells, three cages should provide ample numbers. This means that one afternoon's work should provide sufficient cells for one microarray experiment. It is now abundantly clear that the way forward, as far as cell sorting is concerned, must involve making RNA by the small sample procedure.

Other possible control experiments could involve microarrays on dissociated and sorted cells versus only dissociated cells. This would indicate if the sorting procedure itself has any deleterious effect on the cells.

4.12.3 Possible Applications

Experiments carried out so far have focused on only one proneural gene, *ato*. I chose *ato* as the best one to start with for a number of reasons.

(i) I had carried out microarray experiments on whole embryos to find downstream targets of *ato*. Therefore any genes discovered by cell sorting could be compared to those found from whole embryos.

(ii) There is a large cluster of *ato*-dependent SOPs found in the leg disc.

This cluster is larger than the combined number of *sc*-dependent SOPs found in wing disc. However, especially given the small sample method, it should be possible to obtain sufficient cells from the *sc*-dependent SOPs.

amos is somewhat more complicated as the SOP and PNC cluster enhancers appear to be more tightly linked that those of *sc* and *ato*. Nonetheless, the enhancer I

identified and described in detail in chapter 2, still drives GFP expression in the *amos*-dependent SOPs and, therefore, any population of GFP expressing cells from this line would contain at least a percentage of SOPs. Also GFP expression in the *amos* dependent SOPs of the antennae only occurs following puparium formation making dissection more difficult. Therefore dissociating imaginal disc cells may not be suitable for *amos*.

Although the long-term goal is to compare directly the transcriptional profiles of neural precursor cells i.e. *ato*-dependent SOPs, *sc*-dependent SOPs, *amos*-dependent SOPs with each other, I began by carrying out experiments of the proneural gene *ato*. My initial plan was to compare *ato*-dependent SOPs with non-neural cells, which I had hoped, would identify downstream genes of *ato*. Some of these genes would probably be generic downstream target genes and others would be specific to *ato* and allow the *ato*-dependent SOPs to differentiate specifically into chordotonal organs. These genes could then be compared with those obtained from microarrays on whole embryos using loss-of-function and gain-of-function of *ato* compared to wild type. The *ato* FCO enhancer used in the cell sorting experiments was cloned into a Gal4 vector. Subsequently, the fragment was cloned into a GFP vector by Petra zur Lage and if imaginal disc FACS experiments were to be pursued, this line would be more suitable to use, as delays in turning on the GFP due to the Gal4 would be eliminated.

As outlined above, it may only be possible to compare *ato* specific SOPs with *sc* specific SOPs for imaginal discs. However, all three proneural genes appear suitable for use in embryonic cell dissociation. I have carried out preliminary experiments using the embryonic *ato* enhancer. There are similar enhancers for *sc*

and *amos* which would be suitable to use in these experiments. The *amos* embryonic enhancer gives expression in the *amos* dependent SOPs of the embryo and also in some proneural cluster cells (chapter 2). The *sc* embryonic enhancer gives expression in the *sc* dependent cells of the embryo. Direct comparison between the neural precursor cells of each proneural gene should identify differentially expressed downstream genes. The *amos* and *sc* embryonic enhancers were cloned into a GFP vector rather than a Gal4 one, so that there is less of a delay in switching on the GFP. It would be beneficial to clone the *ato* embryonic enhancer into a GFP vector. This would assist in identifying the correct time period to use and also allow for more accurate comparisons with the other proneural genes.

There is an alternative to using the endogenous enhancers. This would involve utilising fly lines that contain GFP under the control of multimers of *sc* and *ato* specific E-boxes (Powell *et al.*, submitted). These constructs consist of multimerised 20 base pair regions including the E-box and its flanking sequence. The *sc* specific E-box multimer drives GFP expression in the embryo in the *sc* dependent SOPs. The *ato* specific E-box drives expression in the *ato* dependent SOPs. There does not appear to be any proneural cluster expression and so these lines may be more suitable to use in cell sorting experiments, subject to the disadvantage however that there is a delay in the detectability of the GFP. This, although always an important point, becomes crucial when dealing with enhancers that give expression solely in SOPs as SOP formation is highly dynamic. Expression of the proneural gene may be switched off in the SOP before GFP is detectable. Therefore, the time period would need to be more highly defined than that used for a SOP and PNC enhancer, given the extremely short period of time that SOPs are present in the embryo. While we do

not currently have such a reporter line for *amos*, I identified a candidate E-box in the *amos* enhancer region and a multimer of this site may give expression in the *amos*-dependent SOPs.

The above experiments involve comparing SOPs of different proneural genes. If this comparison proves successful, then other future experiments could be designed to compare more subtle SOP distinctions, involving subsets of SOPs. One such experiment could involve comparing recruited versus non-recruited chordotonal organs in the embryo. There are two multimer lines for *ato,* one consisting solely of the six copies of the E-box and the other consisting of six of the E-box and six copies of the Ets site. The Ets site is where Pointed binds and is thought to be responsible for the iterative recruitment of chordotonal organs (zur Lage *et al.*, In prep). There are eight chordotonal organs found in each segment, five of which are formed initially, the remaining three being recruited. Expression of GFP is only found in the original five with the E-box multimer, whereas with the E-box–Ets multimer GFP expression is found in all eight. This means that it should be possible to carry out a microarray on cells sorted from the E-box multimer, compared with cells from the E-box-Ets multimer i.e. *ato*-dependent SOPs versus *ato*-dependent recruited SOPs.

Another possible application could involve using the *amos* embryonic enhancers. *amos* is expressed in the head region as well as in the thorax/abdomen. The upstream region between *amos* and the next closest gene when cloned into a GFP vector supports GFP expression in the *amos*-dependent cells of the embryo in the head and trunk region. Smaller fragments within this region support GFP specifically in the head or trunk. Therefore, embryos expressing GFP specifically in

either the head or the trunk could be dissociated and sorted. This would give two separate GFP positive populations of cells. The cells from the head region are potentially olfactory and those from the trunk become the multidendritic neurons. The transcriptional profiles of these cells could be used to determine which downstream genes are activated by *amos* in the head region when compared to the trunk. This would then shed light on the mechanism whereby *amos* regulates both the formation of olfactory neurons and multidendritic neurons in the embryo.

A third possible experiment, could involve imaginal discs rather than embryos. *ato* is responsible for the chordotonal organs found in the antennae and in the leg disc. The chordotonal organs in the antennae are involved in hearing and those in the leg disc are internal stretch receptors. A comparison between the antennal chordotonal SOPs and the leg chordotonal SOPs, would reveal the similarities and differences between the subsets of *ato* controlled genes, that confer either antennal or leg chordotonal specificity.

Possible experiments described so far involve comparing either SOPs of different proneural genes or subsets of SOPs at a particular time point in development. Another application could be detailed time courses of SOPs from birth to differentiation. This would identify the range of genes involved in specifying SOPs.

4.13 Conclusions

It is possible to use the proneural gene enhancers themselves to drive GFP expression in the SOPs of the imaginal discs and embryos. I have dissociated cells

either from imaginal discs or embryos and have sorted these cells by virtue of GFP expression. This process results in a population of neural precursor cells. Given the recent advances in small sample target labelling, I am confident that sufficient RNA for microarray analysis can be obtained from very few cells and that the transcripts amplified will be a true representation of those found in the sample. Microarray analysis could then be carried out to identify the downstream target genes of each proneural protein. Using specific populations of neural precursor cells as opposed to whole embryos greatly increases the chances of finding downstream target genes of the proneural genes, especially those that are expressed at a low level.

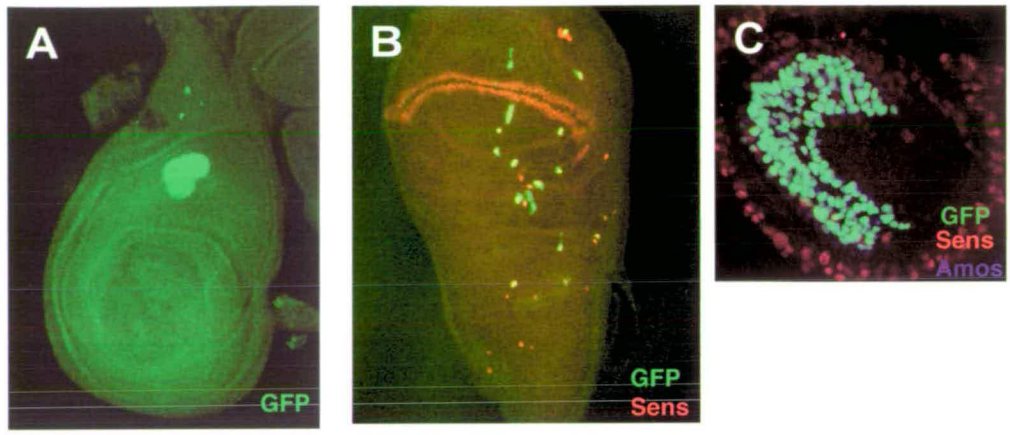


FIG. 4.4.2 GFP expression in imaginal discs. (A) Confocal image of 3rd instar larvae leg disc, *atoGal4, UAS GFP*. It is not necessary to use an antibody to visualise GFP driven by proneural enhancers in antennal discs. GFP is expressed in the SOPs (arrowhead) of the femoral chordotonal organ (FCO) (Picture courtesy of Petra zur Lage.) (B) Confocal image of 3rd instar larva wing disc, *scGal4, UAS GFP*, stained with an antibody to detect Sens protein in red. Sens marks all SOPs. GFP is expressed in the SOPs of the external sense organs (arrowhead). (C) Confocal image of antennal disc 8 h APF, *amos-3.6-GFP*, stained with antibodies to detect Amos protein in blue and Sens protein in red. GFP is expressed in the SOPs (arrowhead) and PNCs.

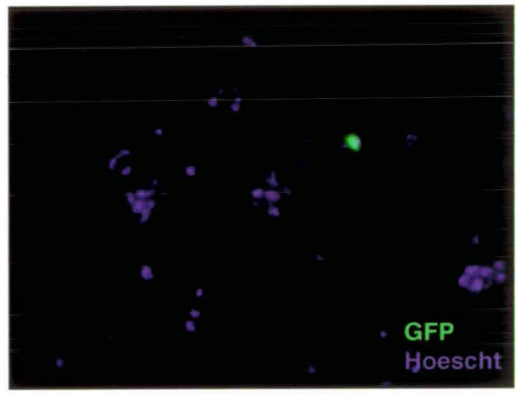


FIG. 4.5.1 GFP-expressing primary imaginal disc cells. Image of cells dissociated from imaginal discs of a *lzGal4, UAS GFP* line using dispase and trypsin (Currie method) and stained with Hoescht in blue. It is not necessary to use an antibody to visualise the GFP.

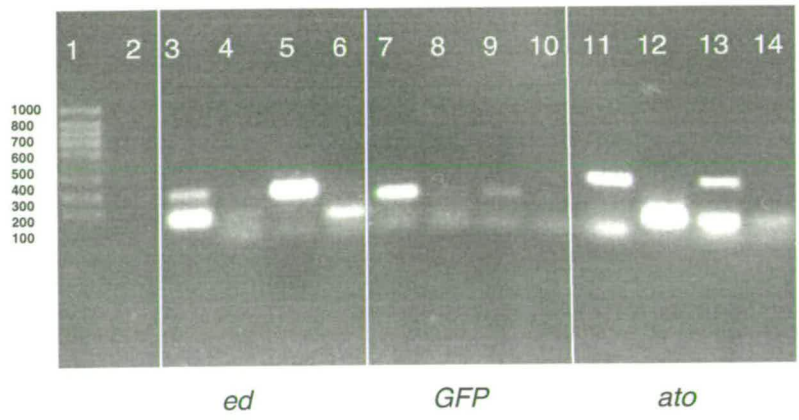


FIG 4.6.1 RT-PCR using RNA extracted from intact leg discs using Trizol. Lane 1, 1 kb ladder. Lane 2, empty. Lane 3, 1 μl of RT reaction containing superscript RT, *ed* primers. Lane 4, 1 μl of RT reaction lacking RT, *ed* primers. Lanes 5 and 6, as lanes 3 and 4, except 0.5 μl of RT reaction. Lane 7, 2μl of RT reaction containing superscript RT, GFP primers. Lane 8, 2 μl of RT reaction lacking RT,GFP primers. Lane 9 and 10, as lanes 7 and 8, except 1.0 μl of RT reaction. Lanes 11-14, as 7-10, except *ato* primers.

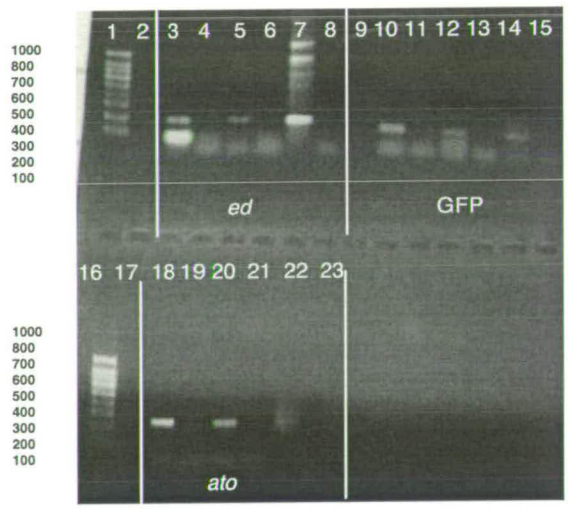


FIG 4.6.2 RT-PCR using RNA extracted from dissociated leg discs using Trizol. Lane 1/17, 1 kb ladder. Lane 2/9/18, empty. Lane 3, 1 μl of RT reaction containing superscript RT, *ed* primers. Lane 4, 1 μl of RT reaction lacking RT, *ed* primers. Lanes 5 and 6, as lanes 3 and 4, except 0.5 μl of RT reaction. Lane 7, PCR using gDNA, *ed* primers. Lane 8, PCR no gDNA, *ed* primers. Lane 10, 2μl of RT reaction containing superscript RT, GFP primers. Lane 11, 2 μl of RT reaction lacking RT, *GFP* primers. Lane 12 and 13, as lanes 10 and 11, except 1.0 μl of RT reaction. Lane 14, PCR using gDNA, GFP primers. Lane 15, PCR no gDNA, GFP primers. Lanes 18-23, as 10-15, except *ato* primers.

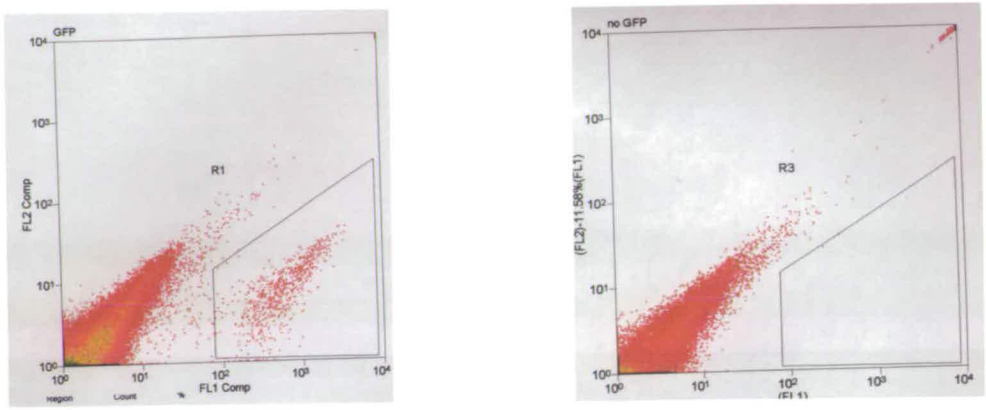


FIG. 4.7.1 Scatterplots showing GFP positive and GFP negative cells. X-axis is GFP fluorescence and Y axis is background fluorescence. (A) Dissociated leg discs from *atoGAl4, UAS GFP*. GFP positive cells are present in the boxed area in the lower right of the graph. (B) Dissociated leg discs from *OrR*. The equivalent region on this plot is almost completely devoid of cells, indicating that the population selected in the case of *atoGal4, UAS GFP* are indeed GFP positive.

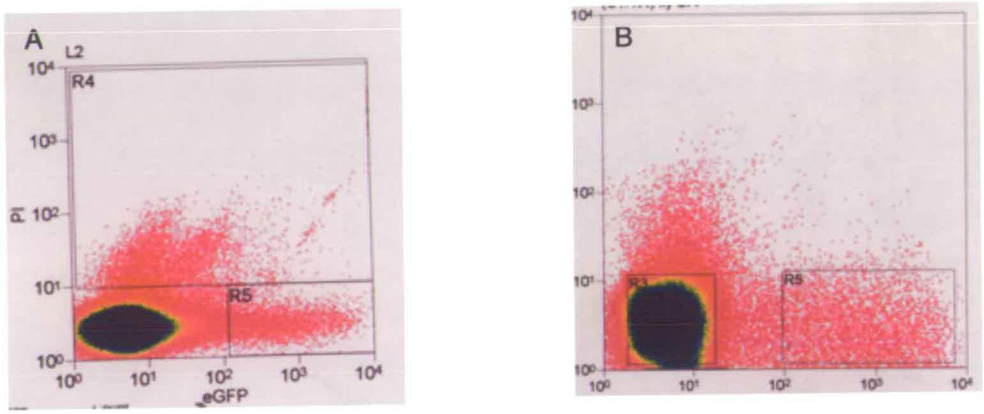


FIG 4.7.2 Propidium iodide staining of sorted cells. (A) Scatterplot of propidium iodide fluorescence versus GFP fluorescence. The cells in R5 are alive (indicated by propidium iodide) and GFP positive. (B) Scatterplot of propidium iodide fluorescence versus GFP fluorescence. Again, the cells in R5 are alive (1.21%).

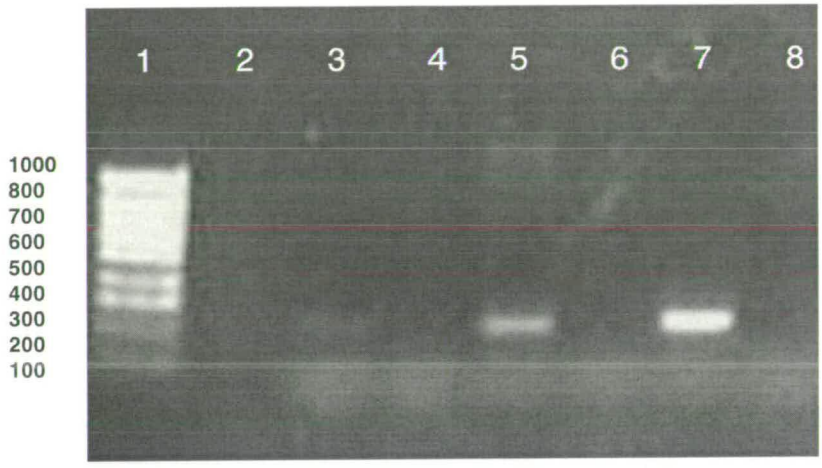
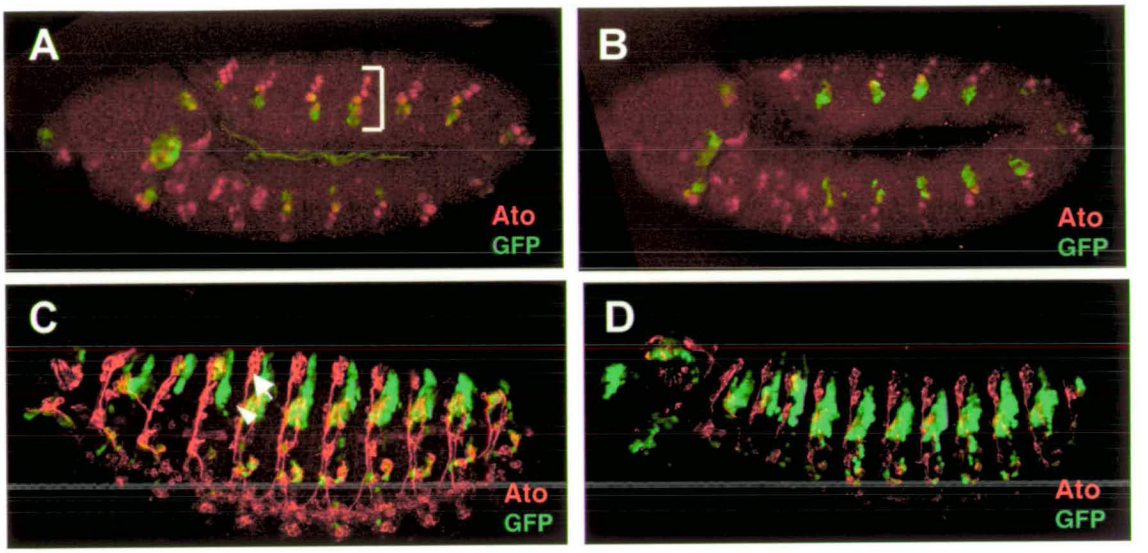


FIG. 4.7.3 RT-PCR on dissociated versus sorted imaginal disc cells. 60,000 sorted cells compared with RT-PCR on RNA made from 60,000 dissociated cells using *histone* primers. Lane 1 1 kb ladder. Lane 2, empty. Lane 3, RT-PCR using RT reaction plus reverse transcriptase on sorted cells. Lane 4, RT-PCR using RT reaction minus reverse transcriptase on sorted cells. Lane 5, RT-PCR using RT reaction plus reverse transcriptase on dissociated cells. Lane 6, RT-PCR using RT reaction minus reverse transcriptase on dissociated cells. Lane 7, PCR, gDNA. Lane 8 PCR, minus gDNA..



4.8.1 *atoGal4, UAS GFP* embryos. (A, B) Confocal images of 6.5-7.5 hours AEL embryos stained with antibodies to detect Ato protein in red and GFP protein in green. Ato protein is present in four cells in each abdominal segment (Bracket). *ato* expression has been switched off in the P cell. GFP is present in the P cells and overlaps with Ato protein in one or two of the more dorsal *ato*-dependent SOPs. GFP is also present in some PNC cells. (C,D) Confocal image of late embryos stained with antibodies to detect 22C10 in red and GFP protein in green. 22C10 stains all neurons. GFP overlaps with chordotonal organs (arrowhead) and is also present in some ectodermal cells (arrow).

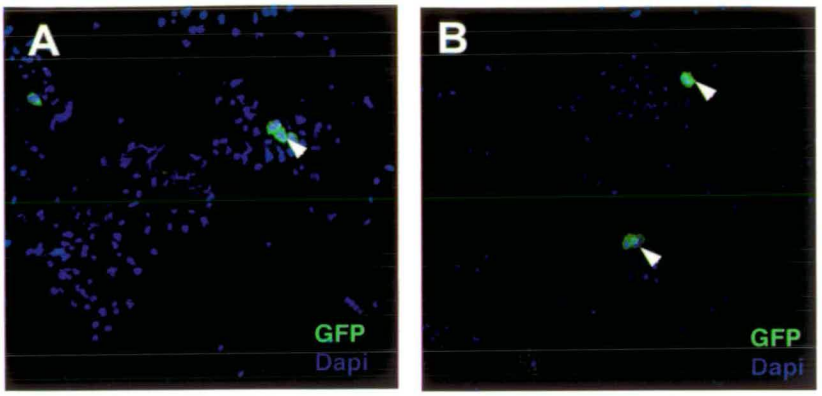


FIG. 4.9.1 GFP expressing dissociated embryonic cells. Confocal image of cells dissociated from embryos (6.5-7.5 h AEL) of genotype *atoGal4, UAS GFP* stained with DAPI to detect DNA. A subpopulation of these cells express GFP (arrowhead). GFP is present in cells of the *ato* dependent neurons and in some ectodermal cells (figure 4.8.1 A/B).

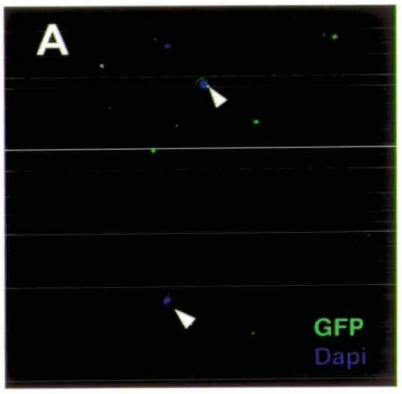


FIG 4.10.3 Sorted embryonic cells. Confocal image of cells dissociated from*atoGal4, UAS GFP* , sorted by virtue of GFP expression and stained with DAPI to detect DNA. GFP positive cells are marked with an arrowhead.

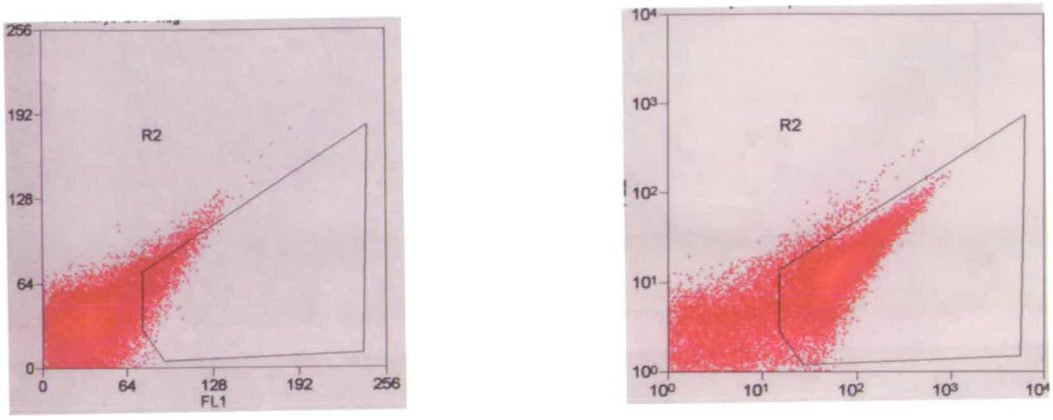


FIG 4.10.1 Scatterplots of GFP positive and GFP negative cells from *OrR* flies and from a line in which nls GFP is expressed ubiquitously. (A) non-GFP expressing line – *OrR* (B) nls GFP is expressed ubiquitously. The boxed region at the lower right of the plot contains GFP expressing cells. The corresponding region for *OrR* should be empty but contains some cells (0.52 %).

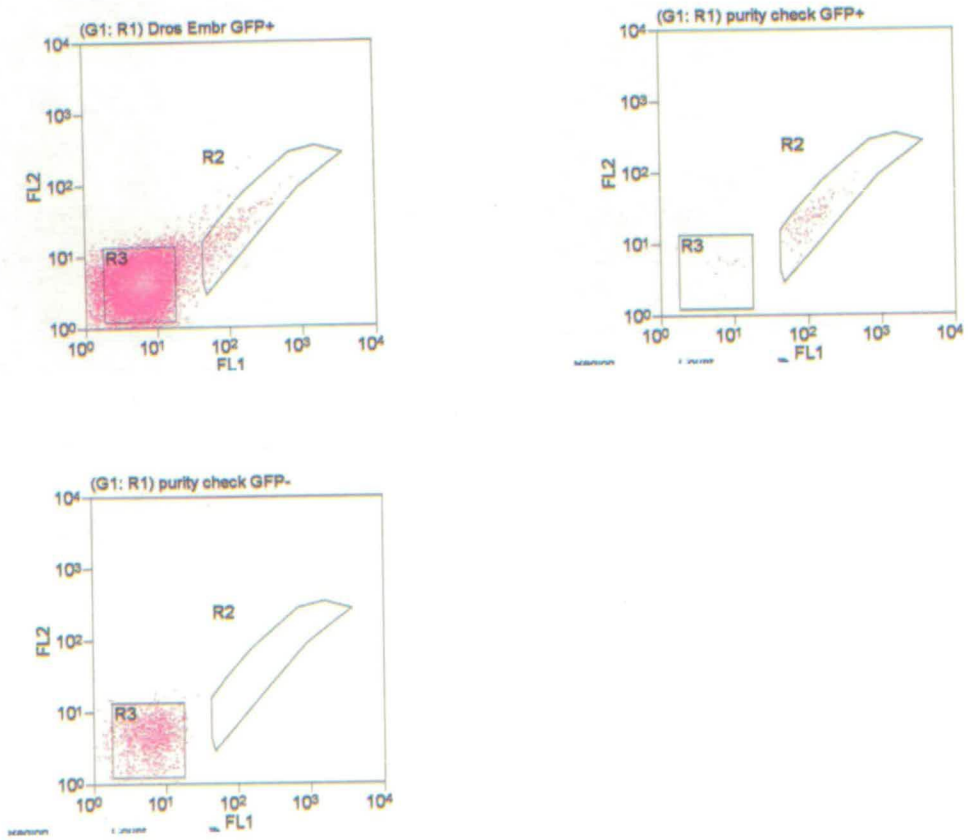


FIG 4.10.2 Scatterplots of GFP positive and negative sorted embryonic cells from *atoGal4, UAS GFP.* (A) Sorted GFP positive and negative embryonic cells. (B) Purity check on previously sorted GFP positive cells, 92.59 % of the cell population is GFP positive. (C) Purity check on previously sorted GFP negative cells.

5

Discussion

5.1 *cis* – regulatory elements required for the transcriptional control of *amos*

My aim in chapter two was to investigate the transcriptional regulation of the proneural gene *amos*. To do this, I cloned fragments upstream of *amos*, into a transformation vector and I assessed the ability of the different fragments to drive reporter gene expression in patterns resembling *amos*.

amos is responsible for two subsets of olfactory sensilla, the sensilla basiconica and sensilla trichodea and it is expressed in the antennal discs from approximately puparium formation onwards, first in a large proneural cluster domain and then in *amos*-dependent SOPs (Goulding *et al.*, 2000; zur Lage *et al.*, 2003). Unusually, expression continues in the PNC following specification of the SOP. *amos* is also involved in embryonic PNS development. In *amos* mutants, the dbd and one of the five dmd neurons are missing (zur Lage *et al.*, 2003). At stage 10 of embryonic development, *amos* is expressed in clusters of cells in the head and in the thoracic and abdominal segments. By stage 11, expression has been switched off in the head (anlage of the olfactory sense organs of the larval antennomaxillary complex), and has been restricted to a single cell per thoracic and abdominal segments.

Subdivision of the *amos* upstream region has shown that separate enhancer elements are responsible for *amos* expression in the antennae and the embryo. In addition, distinct enhancers are responsible for *amos* expression in the head and trunk regions of the embryo. These enhancers may respond to different prepattern factors and/or to autoregulation. Misexpression analysis indicated the possibility of the existence of an autoregulatory component. The presence of E-boxes (binding sites for proneural proteins) in the upstream region indicated that this autoregulation might be direct. I identified a putative Amos/Da binding site that is conserved between *Drosophila melanogaster* and *Drosophila pseudoobscura.* Mutating this site (TT CAAGTG A) led to an overall reduction in GFP expression in the *amos*-dependent PNCs and SOPs of the antennae. This suggests that the two phases of *amos* expression i.e., PNC and SOP are more tightly linked than that of *sc* and *ato.* Unlike *sc* and *ato,* there is no evidence to suggest the existence of a separate enhancer responsible for *amos* expression in SOPs. Furthermore, autoregulation may play a role in both PNC and SOP phases of expression, indicating that the previous delineation of SOP enhancers as being subject to autoregulation, and PNC enhancers as subject to regulation from prepattern factors, may be somewhat artificial.

The putative Amos/Da binding site sequence differs from those identified for Sc/Da and Ato/Da, (although the Ato/Da consensus is based on comparison of only three Ato-regulated target genes). If Amos is indeed binding to this E-box, it may be that Sc, Ato and Amos (although *amos* is *ato* like) prefer and use different sites *in vivo.* The E-boxes responsible for regulation, though sharing many similarities, differ in crucial aspects, lending credence to the idea that different E-boxes underlie different developmental functions. Verification of the sequence TT CAAGTG A as

an Amos/Da binding site will provide a point of reference to investigate how Amos may specifically regulate its putative downstream target genes.

5.2 Microarray analysis to determine the downstream targets of the proneural protein Atonal

The proneural genes are nodal control points in neurogenesis and not only confer neural competence for SOP formation, but also endow SOPs with neural subtype information. This subtype specificity is thought to be due to the regulation by the proneural proteins of differential downstream target genes. Before this possibility can be explored further, the differentially regulated genes must be identified. Therefore, I carried out microarray analysis on the downstream targets of Ato in whole embryos. I initially compared wild-type with mutant with overexpression before concentrating on overexpression analysis. I chose Ato because it is the most suitable of the proneural proteins due to the viability of the mutant (which simplified embryo collections); and also expression of *ato* during embryogenesis should be sufficiently widespread to allow detection of significant changes in gene expression in mutant compared to wild type. Due to the high proportion of unfertilised eggs in the mutant, which will interfere with any comparison I concentrated on overexpression.

I extracted RNA from *scaGal4 x UAS ato* and *scaGal4 x UAS nlsGFP* whole embryos between 6.5 and 7.5 h after egg laying. Microarray analysis was carried out in duplicate at the Sir Henry Wellcome Functional Genomics Facility. Despite the need for more replicates before downstream targets can be identified with certainty, there is some enrichment of interesting genes in the rankings, a number of which

may be true positives. This could be confirmed by *in situ* analysis but the better course of action is first to obtain more microarray data. Future work will, in addition to determining the specific downstream targets of *ato,* involve comparing directly the expression changes induced by misexpression of *sc, ato* and *amos*. This should identify specifically those target genes which are differentially regulated by the different proneural genes rather than co-regulated target genes. *In situ* hybridisation can then be used to verify their involvement in neurogenesis. RNAi and bioinformatic analysis will elucidate their function and help in determining their mode of action.

Results obtained from micorarray experiments so far suggest that it will be possible to identify the more widely expressed downstream targets of the proneural proteins when using the whole embryo approach. However, due to the high proportion of cells in whole embryos that are not specific for SOP formation, it may not be possible to identify genes that, in the context of the whole embryo, are not highly expressed. *TAKR86C* is a case in point. A known downstream target of Ato, it is not detectable on the array. It is expressed only in one of the *ato*-dependent SOPs per hemisegment and this level of expression is not sufficiently high to detect a meaningful change in gene expression. Therefore, I investigated the possibility of using cell sorting to obtain pure populations of neural precursor cells, which could subsequently be used in microarray analysis.

5.3 Exploring cell sorting as a way to compare sense organ precursor transcriptomes

My aim was to develop a system to isolate specific subpopulations of neural precursor cells. This involved constructing or obtaining fly lines that express GFP in the neural precursor cells of the imaginal discs and/or embryos under the control of the proneural gene enhancers. I then dissociated the imaginal discs or embryos and sorted the cells by virtue of GFP expression. Again, I focused on *ato*. This meant that any downstream targets identified could be compared to those identified by whole embryo microarray analysis and also that there is a greater number of GFP positive cells under the control of *ato* (compared to *sc)*. Therefore, it appeared sensible to concentrate on the *atoGal4, UAS GFP* line.

As SOP formation in the embryo is very transient I initially dissociated imaginal disc cells. Given the longer time frame when proneural genes are expressed in the SOPs of the imaginal discs, I thought the advantages were sufficiently high to merit dissociating discs. However, as suspected, it proved difficult to obtain sufficiently high numbers of neural precursor cells from discs and accordingly I turned my attention to dissociating embryos. The advantages are that dissection of embryos is not required and that it is possible to start with a much greater amount of material.

Cell sorting results in a population of neural precursor cells. Recent advances in small sample target labelling mean that sufficient RNA for microarray analysis

can be obtained from very few cells. In addition, the transcripts amplified will be a true representation of those found in the sample. Work up to now has focused on isolating the *ato*-dependent SOPs from both imaginal discs and embryos. The long-term goal is to compare directly the transcriptional profiles of neural precursor cells i.e. *ato*-dependent SOPs, *sc*-dependent SOPs, *amos*-dependent SOPs with each other. As explained, this may not be feasible for imaginal discs but should be possible for embryos. In addition, other possible experiments include comparing subsets of SOPs such as recruited versus non-recruited chordotonal organs in the embryo.

Disc/embryo dissociation followed by cell sorting by virtue of GFP expression, results in a pure population of neural precursor cells. Use of these cells in microarray experiments will help in identifying downstream targets of the proneural proteins, especially those that are not expressed highly at a whole embryo level.

In the coming months I intend to compare directly the expression changes induced by misexpression of *sc, ato* and *amos* in whole embryos. In addition, Sebastian Cachero, a new Phd student, will continue to explore the cell sorting approach, concentrating initially on embryonic cell sorting of GFP expressing *ato*-dependent SOP cells.

5.4 Summary

This thesis has investigated the *cis*-regulatory elements of the *ato*-like proneural gene *amos* and it has established the existence of discrete enhancer elements responsible for *amos* expression in the antennal disc and the head and trunk

regions of the embryo. In addition, I have identified a putative Amos/Da binding site and shown that

(i) SOP and PNC enhancers may not be as separable as thought previously

(ii) autoregulation may have a role to play in both PNC and SOP enhancers.

Another aspect of this work is concerned with the function of the proneural gene, *ato*. I carried out microarray analysis on RNA extracted from whole embryos to determine the downstream targets of Ato. Preliminary results indicate that this approach is capable of identifying more highly expressed genes but it may not be sufficiently sensitive to identify those that are weakly expressed in the context of the whole embryo. Therefore, I investigated the possibility of using cell sorting to obtain pure populations of neural precursor cells which could then be used in microarray analysis. Advances in the small sample labelling procedure suggest that this approach is highly suited to the challenge of identifying downstream targets of the proneural proteins. Identification of these targets will contribute significantly to what is known about the proneural proteins and allow their mechanisms of specificity to be explored.

6

Material and Methods

6.1 Molecular Biology

6.1.1 Preparation of genomic DNA from adult flies

Insect genomic DNA was prepared using the following procedure. 25 flies were frozen in 200µl lysis buffer (100mM Tris-HCl pH 9, 100mM EDTA pH 8, 1% SDS) at –20 °C, thawed and homogenised. A further 200µl of lysis buffer was added and incubated at 70°C for 30 minutes. 150µl of 8M potassium acetate was added, thoroughly mixed and the tube was incubated on ice for 20 minutes. The tube was spun at 14,000 rpm, for 20 minutes at 4 °C. The supernatent was drawn off, split equally between two tubes and 0.9ml of isopropanol added. It was then spun for 5 to 10 minutes, following which the pellet was washed in 70% ethanol and allowed to dry. The pellets were resuspended in 50µl of TE (10mM Tris-HCl, 1mM EDTA, adjusted to pH8) and pooled together. Further purification was carried out by phenol-chloroform extraction. Genomic DNA was precipitated with sodium acetate and ethanol. The solution was mixed and left at –20°C for at least an hour or overnight. The tube was spun at 14,000 rpm for 10-15 minutes, the supernatent discarded and the pellet washed and dried. The pellet was resuspended in 50µl TE.

6.1.2 Preparation of plasmid DNA

Plasmid DNA was prepared using commercial spin columns from Qiagen according to the manufacturer's instructions. For DNA injection of embryos to generate transformants, the following bulk prep procedure was used.

6.1.3 Plasmid bulk prep

Bacterial cultures were transferred to 50ml Falcon tubes and centrifuged at 4,500 rpm for 15 minutes. The pellets were drained thoroughly and resuspended carefully using a pastette in 2ml of solution I (50mM Glucose, 25mM Tris pH 8, 10mM EDTA, 5mg/ml lysozyme, added just before use), per 50ml of culture and left at room temperature for 10 minutes. 4ml of Solution II (0.2 M NaOH, 1% SDS - Made just before use from 10M NaOH and 10% SDS stocks) was added and mixed thoroughly but not vigorously. After 10 minutes on ice with occasional mixing 3ml of Solution III (3M KOAc / 1.3M HCOOH) were added. The solution was mixed quickly and thoroughly and placed on ice for 15 minutes. The mixture was centrifuged at 700 rpm for 15 minutes. The mixture was transferred to a clean tube avoiding transfer of any precipitate. 0.6 volumes of 100% isopropanol was added and the solution was left standing for 5 minutes followed by centrifugation at 4,000 rpm for 10 minutes. The supernatent was discarded and the pellet was rinsed with approx 2ml 70% ethanol. The tube was wiped out and the still wet pellet dissolved in 1ml TE. The DNA solution was transferred to eppendorfs and placed on ice for 5-10 minutes. An equal volume of cold 5M LiCl (stored at –20°C) was added and placed on ice for 5 minutes. The tubes were centrifuged for 5 minutes at 14,000 rpm. The supernatant was transferred to clean eppendorf tubes (on ice) and an equal

volume of isopropanol was added. The tubes were left on ice for 10 minutes and then spun for 5 minutes. The supernatent was discarded and the pellets airdryed. They were then redissolved in a total of 300 µl TE. 1µl Dnase-free Rnase (10mg/ml stock) was added and the mixture incubated at 37 °C for 30 minutes and then transferred to ice. An equal volume of PEG/NaCl (15% PEG, 1.6M NaCl) was added, and the tube left on ice for 5 minutes before spinning for 5 minutes. The supernatent was discarded and the pellet was resuspended in 300 µl TE. A PhOH/$CHCl_3$ extraction was carried out (add equal volume of PhOH/$ChCl_3$, vortex, spin, remove top aqueous phase) followed by a $CHCl_3$ extraction. The aqueous phase was transferred to a clean tube(s). The DNA was precipitated by addition of 1/20 vol 3M NaOAc (pH 5.6-6) and 2 vol 100% EtOH. This was thoroughly mixed and a large precipitate was usually seen. The tubes were centrifuged for 5 minutes, the pellets washed with 70% EtOH, air dried and resuspended in 300 µl TE.

6.1.4 RNA extraction

RNA was extracted using either TRIzol Reagent (Invitrogen), a GenElute Mammalian Total RNA (Sigma) kit or RNAeasy kit from Qiagen, according to the manufacturers instructions. In the case of Trizol Reagent the initial volume used was 0.5ml of TRIzol and the RNA resuspended in 20µl of nuclease-free water. Cells were lysed or tissues homogenised in 500 µl of lysis soulution/2-ME mixture when using the GenElute kit and 600 µl when using the Qiagen kit. If the RNA was to be used in RT-PCR reactions it was then treated with Dnase (Ambion) to remove contaminating DNA from RNA preparations. A Cells to cDNA kit was tested

(Ambion) to avoid the need to extract RNA before cDNA preparation but this proved unsuccessful.

6.1.5 Separation of DNA fragments by gel electrophoresis

DNA was analysed using standard gel electrophoresis. Usually 0.8% agarose in 1xTAE containing 0.5 μg/ml EtBr solutions were used. For separation of low molecular weight fragments, 2% agarose gels were used. DNA was mixed with loading dye before running on the gel. Gels were run at 80V or 100V depending on the size of the gel tank for varying times.

6.1.6 Estimation of nucleic acid concentration

The concentration of nucleic acid solutions was determined either by gel electophoresis with reference to known molecular weight standards (Bioline) or by spectrophotometry. Absorbance of a DNA/RNA sample was read at 260nm. To assess RNA purity, an A_{260}:A_{280} ratio was utilised.

6.1.7 Polymerase Chain Reaction (PCR)

PCR was used to amplify fragments for subcloning and to check for specific sequences in isolated DNA and RNA. In all cases primers were initially tested using appropriate positive control template DNA. Roche and Stratagene *Taq* and their supplied buffers were used. A standard magnesium chloride concentration of 1.5 mM was used. Between 30 and 50 pmol of each primer was used per 50 μl of reaction and 2.5 nmol of dNTPs. Negative controls (without DNA) were performed in parallel for each primer set. The standard PCR program used was an initial

denaturing step 94°C for 2 minutes followed by 30 cycles of 94°C for up to 1 minute, 50-60 °C annealing for up to one minute, extension at 72 °C or 1 minute followed by final extension cycle of 10 minutes, then left at 4 °C. A personal thermal cycler, either Biometra or Thermo Hybaid was used for all PCR reactions and the lid temperature was set to 100 °C. For RT-PCR, superscript II an Rnase H^- Reverse Transcriptase from Invitrogen was used to prepare cDNA followed by polymerase chain reaction.

Primer List

Primers used to amplify *sc* enhancer (EcoRI, BamHI) 356bp
EH 1 GTC GGA TCC GAA CCT TTG GGT TGA GCA AA
EH 2 GTC AAG CTT CCC TCC TTG CCC TTA AGA CT

Primers used to amplify *senseless*
EH 13 CAA CAC TGT GAT AGC CTC
EH 14 GCT CCT GAT AGT CCT GCT

Primers used to amplify *echinoid*
E5-5 CTT GGT ATC GCG AGG GTT CG
E5-3 TAA CAT TCA ATG TGG CAG TGG C

Primers used to amplify *histone2A*
GTG GAA AAG GTG GCA AAG TGA A
ATG CTG GTG ACA ACA AGA A

Primers used to amplify *GFP*
EH5 GCC ATG GAA CAG GTA GTT
EH6 CAC TGG AGT TGT CCC AAT

Primers used to amplify *atonal*
EH 11 AAG ACC TCC GAG GAC TTG
EH 12 GCC TCC ATT GTA CAT GGT

Primers used to amplify *amos-A-GFP*
EH 3 GTC CCG CGG GGA GTG CAA CCG GAT TTA ACC
EH 25 CCT AGC GAA AGC GGA GAA TT

Primers used to amplify *amos-B-GFP*
EH 17 AAT TCT CCG CTT TCG CTA GG
EH 18 CGA GGA GTT CGC TGA ATT TC

Primers used to amplify *amos-C-GFP*
EH 26 GAA ATT CAG CGA ACT CCT CG
EH 27 CCC GAT TGC CAA CCT CTT GA

Primers used to amplify *amos-AB-GFP*
EH 3 GTC CCG CGG GGA GTG CAA CCG GAT TTA ACC
EH 18 CGA GGA GTT CGC TGA ATT TC

Primers used to amplify *amos-BC-GFP*
EH 17 AAT TCT CCG CTT TCG CTA GG
EH 27 CCC GAT TGC CAA CCT CTT GA

Primers used for site directed mutagenesis
E-box Amos 1 CCG AGA TTA TCA TGG ATC CAC CTA TTG TTG GTC AG
E-box Amos 1 CTG ACC AAT AGG TGG ATC CAT GAT AAT CTC GG

E-box Amos 2 CTT TCG ATG CGG TGG ATC CAA ATG GGG CTT GAT CC
E-box Amos 2 GGA TCA AGC CCA CTT TGG ATC CAC CGC ATC GAA AG

6.1.8 Restriction digests

Restriction endonucleases were used according the manufactures instructions. (Roche). To screen plasmid minipreps for the appropriate insert approximately 3 units of enzyme were used per μg of DNA and incubated in the appropriate buffer for up to one hour at 37°C. For digestions of bulk prep DNA reactions were scaled up and left for 2 hours at 37 °C.

6.1.9 5' dephosphorylation

To prevent recircularisation of vector DNA in ligation reactions, vectors were dephosphorylated at 5' ends using calf intestinal phosphatae (CIP) in ligation

reactions. 1µl of CIP was added to DNA in digestion buffer for 30 minutes at 37 °C and a further 1µl of CIP was added for another 30 minutes.

6.1.10 Ligations

T4 DNA ligase (NEB) was used according to the manufacturer's instructions.

6.1.11 Site directed mutagenesis

The QuikChange Site-Directed Mutagenesis kit (Stratagene) was used according to the manufacturer's instructions.

6.1.12 Bacterial culture growth

The medium used for culture of E.coli was autoclaved Luria-Bertani (LB) or XZY broth. Liquid colonies were grown by incubation at 37^{0}C in an orbital shaker, or single colonies were streaked out on solid LB agarose medium and incubated at 37°C overnight. Media were supplied with appropriate nutrients and antibiotics after sterilisation. Ampicillin was added to a concentration of 50µl/ml.

6.1.13 Transformation of *E. coli*

Competent cells used for transformation were either prepared using a $CaCl_2$ procedure (David Prentice) or commercially available cells (Invitrogen) were used for the TA cloning procedure, or for site directed mutagenesis. Commercially available Xl-10 Gold cells were also used in some instances. In these cases the appropriate transformation protocol was followed according to the manufacturers instructions.

10-100ng of DNA in ligation buffer (2-5μl) was added to 100μl of competent cells, which were left on ice or 30 minutes for adsorption. Cells were heat shocked for 45 seconds at 42 °C for DNA uptake and were allowed to recover on ice for 2 minutes. 0.9ml of SOC medium was added and the tubes were incubated at 37 °C for 30 minutes with gentle shaking. 200μl of the transformation reaction were spread on LB plates containing the appropriate antibiotic using a sterile spreader. The remaining cells were centrifuged at 3000rpm for 5 minutes and the pellet re-suspended in 200μl of SOC medium and plated. For blue/white selection, 100μl 100mM IPTG and 100μl 2%X-gal were added to the agar plates prior to plating the transformations. The plates were incubated overnight at 37 °C. Single colonies were picked using a sterile loop and grown overnight in 5ml cultures of LB broth and antibiotic.

6.1.14 DNA Sequencing

A BigDye Dye terminator kit (Perkin Elmer Applied Biosystems) was used according to the manufacturers instructions and in a reaction consisting of 4.0 μl reaction mix, between 250-500 ng of template DNA, 1.6 μl pmol of primer and sterile H_2O to a final volume of 10 μl. Cycle conditions were 31 cycles of melting at 92°C for 30 seconds, annealing at 55°C for 30 seconds and elongation at 72°C for two minutes. 10 μl of dH_20 were added and the reaction analysed on an ABI377 sequencer at ICMB, University of Edinburgh. The sequence was analysed using Gene Jockey II (P.L. Taylor, Biosoft, UK).

6.2 Immunohistochemical procedures

6.2.1 Fixation of embryos and imaginal discs

Embryos were collected for fixation on grape juice plates and removed using water and a paint- brush to a sieve. Embryos were washed to remove yeast, dechorionated in 50% bleach for three minutes, and thoroughly washed to ensure all bleach was removed. They were then fixed in 1.25 ml formaldehyde (37%), 3.75ml PBS (8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4 for 1 litre, adjusted to pH7.4) and 5 ml n-Heptane (Sigma) in a scintillation vial. Embryos were fixed with shaking for 20 minutes. Following shaking, as much of the bottom phase as possible was removed without removing the embryos. Two volumes of methanol were added and the vial shaken vigorously for 30 seconds to devitellinise the embryos. Embryos were allowed to settle at the bottom of the vial, transferred to an eppendorf and rinsed first with methanol to remove residual heptane, and then with PBTx (PBS plus 0.3% Triton X-100, Sigma) and used directly.

Larval and pupal imaginal discs were dissected at room temperature in PBS, (pupae were pinned to watchglass), fixed for five minutes in 4% formaldehyde in PBS, rinsed in PBTx three times, washed 3 x 15 minutes and used immediately.

6.2.2 Immunohistochemistry

Embryos and imaginal discs were blocked for at least two hours in 2% bovine serum albumin (BSA) solution (Sigma) in PBTx at room temperature with gentle

shaking. Primary antibody, at the appropriate concentration in 2% BSA, Normal Goat Serum (NGS, Jackson labs) and PBTx and was then added and samples incubated at 4°C overnight. Table 6.7.1 lists the antibodies used, and their concentrations. Samples were then rinsed and washed 3 x 15 minutes in PBTx. The secondary antibody (flurochrome conjugate) was added in PBTx to a concentration of 1:1000 for 2 hours at hour temperature. The samples were rinsed in PBTx and then washed 3 x 15 minutes. Samples were mounted (dissected first in the case of imaginal discs) in Vectashield mounting medium (Vector labs) on microscope slides sealed with a coverslip and nail varnish. Slides were stored in the dark at 4°C. Confocal images were taken on a Leica TCS-NT microscope, uing Leica TCS-NT image capture software.

Antibody	Concentration	Staining pattern	Reference/Source
Rabbit α Atonal	1: 2000	*ato*-dependent SOPs	(Jarman *et al.*, 1993)
Rabbit α Amos	1: 1000	*amos*-dependent SOPs	(zur Lage *et al.*, 2003)
Guinea pig α Senseless	1: 5000	All SOPs	(Nolo *et al.*, 2000)
Rat α Elav	1:200	Neuron nucleus	Developmental Studies Hybridoma Bank, University of Iowa
Mouse α Prospero	1:200	Support Cells	Developmental Studies Hybridoma Bank, University of Iowa
Mouse α Cut	1: 200	Support Cells	Developmental Studies Hybridoma Bank, University of Iowa
Mouse α 22C10	1: 200	Neuronal cell body and processes	(Zipursky *et al.*, 1984)
Rabbit/Mouse α GFP	1: 1000	N/A	Molecular Probes

Table 6.2.1 Primary antibodies used and their concentrations

6.3 Injection of DNA

Constructs containing the enhancer of interest in either a GAL4 or pHStinger vector (P element vector) were injected into Δ 2-3 flies. The Δ 2-3 is the source of transposase for the attenuated P element vector. DNA is introduced into pre-cellular blastoderm embryos by injection and integrated into the genome by random transposition events. DNA for each construct was prepared using the Bulk Prep method.

Cages of flies were set up and the grape-juice agar plate with yeast paste changed regularly to encourage laying. Plates were collected every hour and the embryos used for injection. The injection procedure was carried out at 18 °C. Embryos were dechorionated for 3 minutes in 50% bleach and then rinsed in H_2O. Embryos were lined up under a microscope along the edge of a piece of agar in one orientation. They were then transferred to a coverslip coated with a film of glue. The coverslip was attached to a microscope slide using a drop of oil and placed in silica gel at room temperature for approx 10 minutes to allow dehydration. Embryos were then covered with series 700 halocarbon oil and injected with the construct of interest. Injected embryos were then covered in series 95 halocarbon oil, left at 18 °C for one day and then allowed to develop at 25 °C. Adult flies were crossed with white eyed flies (w^{1118}) and transformants screened for on the basis of eye colour.

6.4 Cell Dissociation

6.4.1 Imaginal Discs

Ten or more third instar larvae were dissected in PBS or Graces Medium to remove imaginal discs and/or brains. A number of different methods of dissociating were explored.

Fehon and Schubiger

The first method attempted was the Fehon method. Larvae were dissected in Schneider's medium. Discs were moved to an eppendorf containing fresh medium and 2mg/ml of collagenase was added. At three ten minutes intervals the mixture was gently pipetted (using a 200 µl eppendorf pipette) 20 times. Discs were rinsed

three times in Schneider's medium. 50 µl of 0.35 M citric acid was added and the mixture pipetted 12 times. The entire mixture was transferred to a solution containing 0.2ml of Schneider's medium plus 0.18 ml of 0.3 M Tris base. Any large pieces remaining were allowed to settle out for 8-10 minutes and then the supernatent was transferred to a new tube, centrifuged at 150 g at 4 °C and the resulting pellet resuspended in 100 µl of medium.

Currie

The Currie method is as follows. Larvae were dissected in Shields and Sang medium. The medium was removed and PBS containing 2 mg/ml of dispase and 4mM EDTA was added. Discs were gently shaken at room temperature for 45 minutes. Bacto trypsin was added to a final concentration of 0.1% and left for a further ten minutes. The solution was transferred to an eppendorf and further dissociated by brief shaking on a vortex mixer. It was then centrifuged for four minutes at 150g at 4°C to pellet single cells and small clumps. The supernatent was removed replaced by 100 µl of medium and pipetted 50 times.

Wyss

Larvae were dissected and the discs transferred to 1 ml of medium. The resulting mixture was transferred to an eppendorf. 50 µl of a 5 mg/ml solution of protease VIII in water was added. The tube was incubated for 1 h at 25 ° C followed by centrifugation for one minute at 1,000g. The pellet was resuspended in 100 µl and the cells were pipetted 40 times to dissociate any remaining clumps.

Neufeld

Larvae were dissected and the discs transferred to a glass dish containing 500µl of 9x Trypsin-EDTA and 1X PBS. The discs were dissociated at room

temperature for two hours with gentle agitation. They were then transferred to an eppendorf, centrifuged and the pellet resuspendend in 100 μl of Shields and Sang medium by pipetting 40 times.

10 μl of the suspension was placed on a haemacytometer and the number of cells present were counted. If the cells were to be used for FACS analysis, the solution was filtered through a 40 μm mesh.

6.4.2 Embryos

Embryos of the appropriate age were collected from a red wine agar plate, dechorionated for 2 minutes and washed thoroughly. They were then transferred to a 7 ml Dounce homogeniser and dissociated in Shields and Sang culture medium using 24 strokes of type A (loose) pestle. The resulting suspension was strained through a filter and centrifuged for 5 minutes at 3,000g. The cells were resuspended in 100 μl of Shields and Sang medium by pippetting.

6.5 FACS Analysis

Cell sorting on the basis of GFP expression was carried out using a Cytomation Moflo cell sorter installed in the Institute for Stem Cell Research.

6.6 Microarray analysis

Microarray analysis was carried out at the Scottish Centre for Genomic Technology at Little France, Edinburgh or the Sir Henry Wellcome Functional Genomics Facility. RNA was extracted and delivered on dry ice to the centre. The quality of the RNA was assessed using an Agilent Bioanalyzer according to the manufacturers instructions. cDNA, cRNA production, cRNA labelling and hybridisation to arrays was carried out by personnel at the centre using the Affymetrix standard procedure.

6.6.1 Statistics

Statistical analysis was carried out using Excel. See CD.

6.7 *Drosophila* strains

Fly stocks were maintained on standard cornmeal-agar medium ("Dundee Food") prepared by Swann media kitchen staff at 18°C or room temperature. *OreganR* flies were used as the wildtype strain throughout. Fly stocks are listed in appendices A and B.

Appendix

A. General Fly Stocks

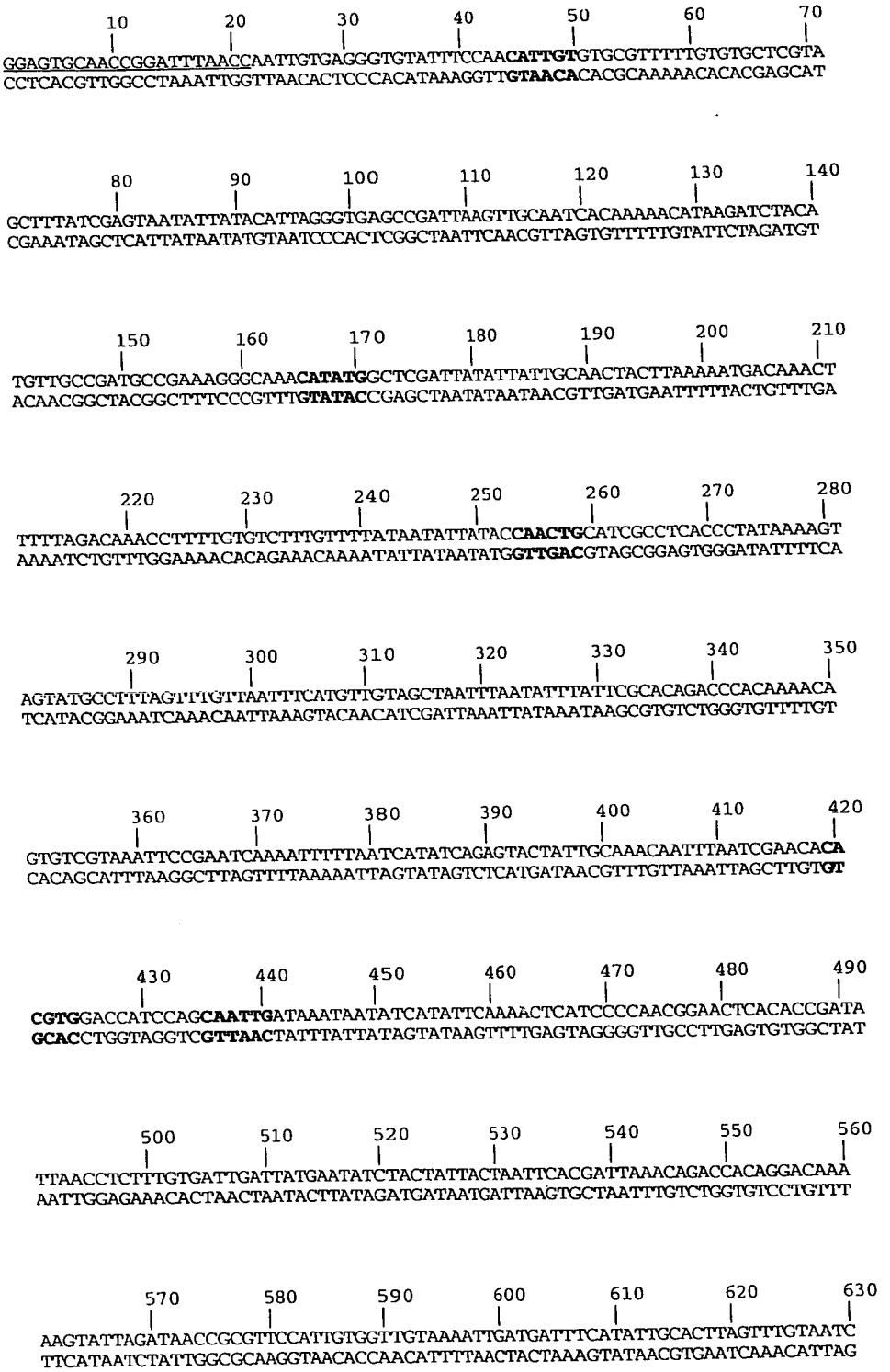
Genotype	Nature of allele	Source/reference
yw;109-68Gal4	Gal4 driver	(Jarman and Ahmed, 1998)
pPTGal4	Gal 4 driver	(Sharma *et al.*, 2002)
Δ 2-3	Source of transposase	Gift of Hiro Ohkuru
scaGal4	Gal4 driver	(Egger *et al.*, 2002)
yw; Pin/Cyo	Visible, dominant, balancer	Bloomington
w; Ly/TM3	Visible, dominant, balancer	Bloomington
UAS ato	Construct	(Jarman *et al.*, 1993)
UAS amos	Construct	(Goulding *et al.*, 2000)
UAS GFP	Construct	Bloomington
UAS nlsGFP	Construct	Bloomington
ato^{1}	Null	(Jarman *et al.*, 1993)

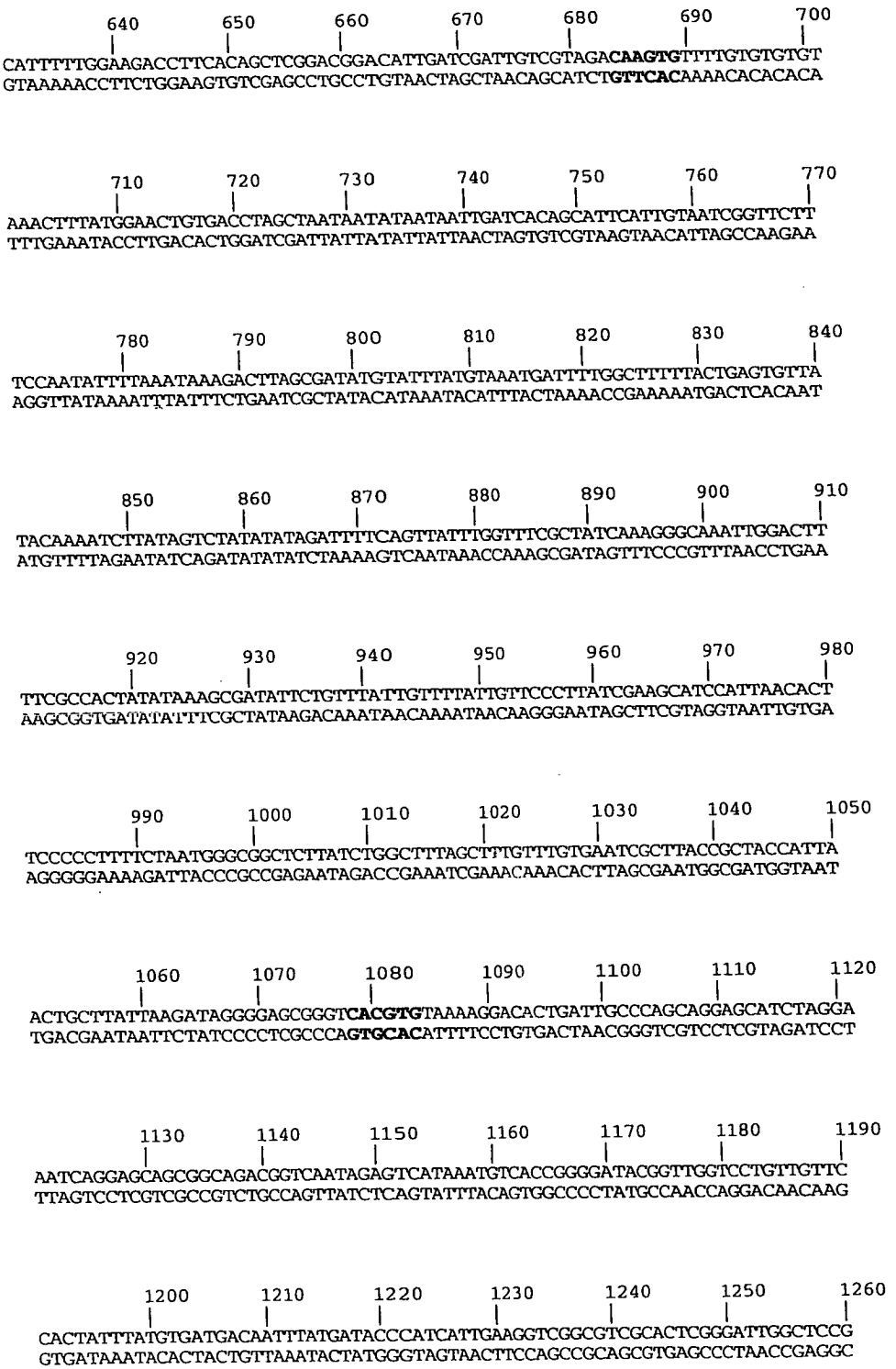
Appendix B: *amos* regulatory region pH Stinger lines

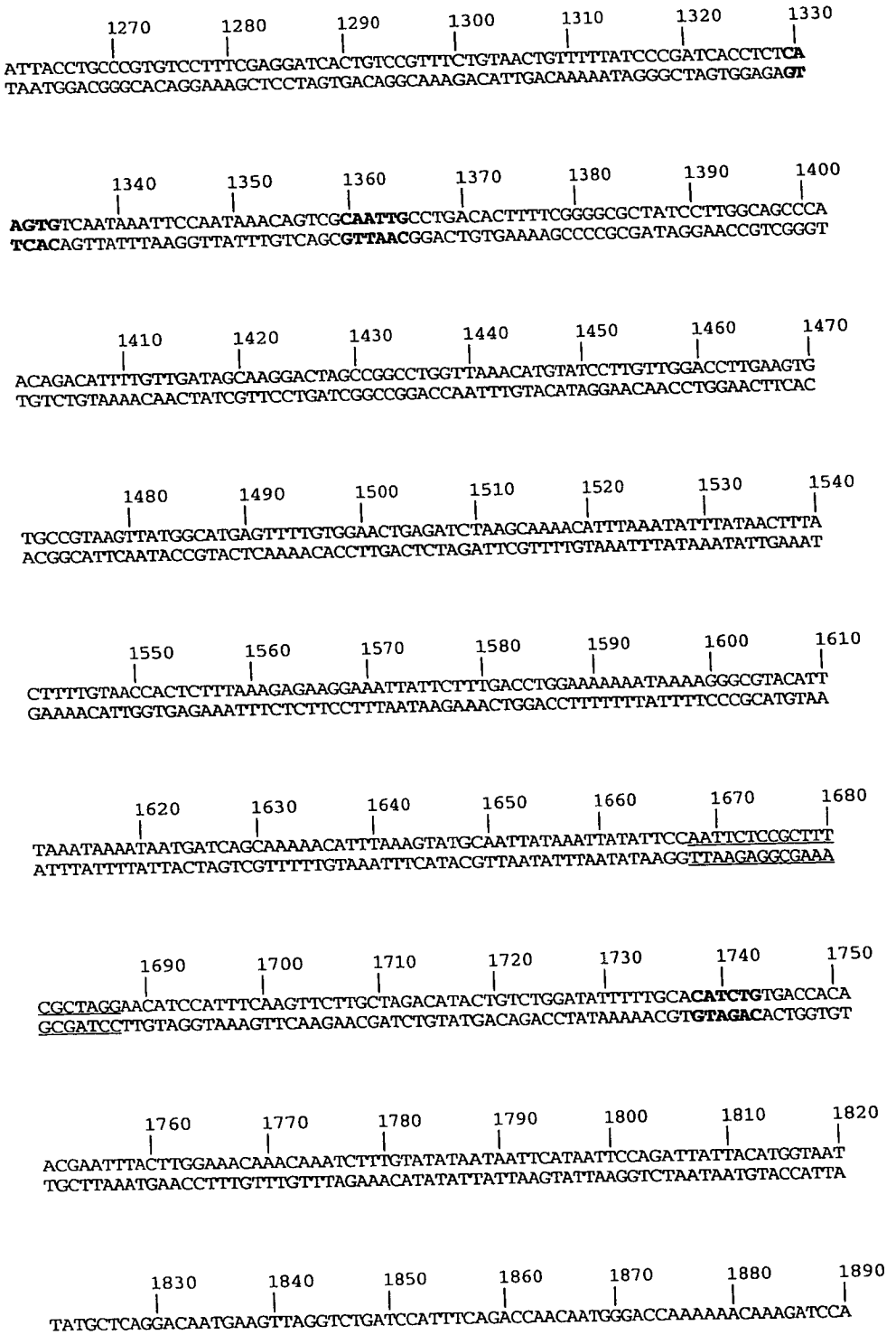
	Chromosome
3.6 1	II
3.6 2	III
3.6 3	X
3.6 4	X
3.6 5	II
3.6 6	II
3.6 M1	X
3.6 M 2	II
A	II
B 1	X
B 2	II
C	III
AB 1	III
AB 2	III
AB 3	III
AB 4	II
AB 5	X
BC 4	III
C Mutated E-box	III

All lines are homozygous viable except for C Mutated E-box

Appendix C: Sequence of *amos* regulatory region




640 650 660 670 680 690 700
CATTTTTGGAAGACCTTCACAGCTCGGACGGACATTGATCGATTGTCGTAGACAAGTGTTTTGTGTGTGT
GTAAAAACCTTCTGGAAGTGTCGAGCCTGCCTGTAACTAGCTAACAGCATCTGTTCACAAAACACACACA
710 720 730 740 750 760 770
AAACTTTATGGAACTGTGACCTAGCTAATAATATAATAATTGATCACAGCATTCATTGTAATCGGTTCTT
TTTGAAATACCTTGACACTGGATCGATTATTATATTATTAACTAGTGTCGTAAGTAACATTAGCCAAGAA
780 790 800 810 820 830 840
TCCAATATTTTAAATAAAGACTTAGCGATATGTATTTATGTAAATGATTTTGGCTTTTTACTGAGTGTTA
AGGTTATAAAATTTATTTCTGAATCGCTATACATAAATACATTTACTAAAACCGAAAAATGACTCACAAT
850 860 870 880 890 900 910
TACAAAATCTTATAGTCTATATATAGATTTTCAGTTATTTGGTTTCGCTATCAAAGGGCAAATTGGACTT
ATGTTTTAGAATATCAGATATATATCTAAAAGTCAATAAACCAAAGCGATAGTTTCCCGTTTAACCTGAA
920 930 940 950 960 970 980
TTCGCCACTATATAAAGCGATATTCTGTTTATTGTTTTATTGTTCCCTTATCGAAGCATCCATTAACACT
AAGCGGTGATATATTTCGCTATAAGACAAATAACAAAATAACAAGGGAATAGCTTCGTAGGTAATTGTGA
990 1000 1010 1020 1030 1040 1050
TCCCCCTTTTCTAATGGGCGGCTCTTATCTGGCTTTAGCTTTGTTTGTGAATCGCTTACCGCTACCATTA
AGGGGGAAAAGATTACCCGCCGAGAATAGACCGAAATCGAAACAAACACTTAGCGAATGGCGATGGTAAT
1060 1070 1080 1090 1100 1110 1120
ACTGCTTATTAAGATAGGGGAGCGGGTCACGTGTAAAAGGACACTGATTGCCCAGCAGGAGCATCTAGGA
TGACGAATAATTCTATCCCCTCGCCCAGTGCACATTTTCCTGTGACTAACGGGTCGTCCTCGTAGATCCT
1130 1140 1150 1160 1170 1180 1190
AATCAGGAGCAGCGGCAGACGGTCAATAGAGTCATAAATGTCACCGGGGATACGGTTGGTCCTGTTGTTC
TTAGTCCTCGTCGCCGTCTGCCAGTTATCTCAGTATTTACAGTGGCCCCTATGCCAACCAGGACAACAAG
1200 1210 1220 1230 1240 1250 1260
CACTATTTATGTGATGACAATTTATGATACCCATCATTGAAGGTCGGCGTCGCACTCGGGATTGGCTCCG
GTGATAAATACACTACTGTTAAATACTATGGGTAGTAACTTCCAGCCGCAGCGTGAGCCCTAACCGAGGC


1270 1280 1290 1300 1310 1320 1330
ATTACCTGCCCGTGTCCTTTCGAGGATCACTGTCCGTTTCTGTAACTGTTTTTATCCCGATCACCTCTCA
TAATGGACGGGCACAGGAAAGCTCCTAGTGACAGGCAAAGACATTGACAAAAATAGGGCTAGTGGAGAGT
1340 1350 1360 1370 1380 1390 1400
AGTGTCAATAAATTCCAATAAACAGTCGCAATTGCCTGACACTTTTCGGGGCGCTATCCTTGGCAGCCCA
TCACAGTTATTTAAGGTTATTTGTCAGCGTTAACGGACTGTGAAAAGCCCCGCGATAGGAACCGTCGGGT
1410 1420 1430 1440 1450 1460 1470
ACAGACATTTTGTTGATAGCAAGGACTAGCCGGCCTGGTTAAACATGTATCCTTGTTGGACCTTGAAGTG
TGTCTGTAAAACAACTATCGTTCCTGATCGGCCGGACCAATTTGTACATAGGAACAACCTGGAACTTCAC
1480 1490 1500 1510 1520 1530 1540
TGCCGTAAGTTATGGCATGAGTTTTGTGGAACTGAGATCTAAGCAAAACATTTAAATATTTATAACTTTA
ACGGCATTCAATACCGTACTCAAAACACCTTGACTCTAGATTCGTTTTGTAAATTTATAAATATTGAAAT
1550 1560 1570 1580 1590 1600 1610
CTTTTGTAACCACTCTTTAAAGAGAAGGAAATTATTCTTTGACCTGGAAAAAAATAAAAGGGCGTACATT
GAAAACATTGGTGAGAAATTTCTCTTCCTTTAATAAGAAACTGGACCTTTTTTTATTTTCCCGCATGTAA
1620 1630 1640 1650 1660 1670 1680
TAAATAAAATAATGATCAGCAAAAACATTTAAAGTATGCAATTATAAATTATATTCCAATTCTCCGCTTT
ATTTATTTTATTACTAGTCGTTTTTGTAAATTTCATACGTTAATATTTAATATAAGGTTAAGAGGCGAAA
1690 1700 1710 1720 1730 1740 1750
CGCTAGGAACATCCATTTCAAGTTCTTGCTAGACATACTGTCTGGATATTTTTGCACATCTGTGACCACA
GCGATCCTTGTAGGTAAAGTTCAAGAACGATCTGTATGACAGACCTATAAAAACGTGTAGACACTGGTGT
1760 1770 1780 1790 1800 1810 1820
ACGAATTTACTTGGAAACAAACAAATCTTTGTATATAATAATTCATAATTCCAGATTATTACATGGTAAT
TGCTTAAATGAACCTTTGTTTGTTTAGAAACATATATTATTAAGTATTAAGGTCTAATAATGTACCATTA
1830 1840 1850 1860 1870 1880 1890
TATGCTCAGGACAATGAAGTTAGGTCTGATCCATTTCAGACCAACAATGGGACCAAAAAACAAAGATCCA

ATACGAGTCCTGTTACTTCAATCCAGACTAGGTAAAGTCTGGTTGTTACCCTGGTTTTTTGTTTCTAGGT

1900 1910 1920 1930 1940 1950 1960

TTGTCAGGCATTGCAATACAAAGAGTGGAATCAA**CAAGTG**TCGGGCGAGAAGAGATTCCCATGGAATTCC
AACAGTCCGTAACGTTATGTTTCTCACCTTAGTT**GTTCAC**AGCCCGCTCTTCTCTAAGGGTACCTTAAGG

1970 1980 1990 2000 2010 2020 2030

TGGTGACTCCGGGACTCTGACTTCTAATCTCATTTTCGCAATTTAGCGCCAACAACTAATCCGCAAAGCG
ACCACTGAGGCCCTGAGACTGAAGATTAGAGTAAAAGCGTTAAATCGCGGTTGTTGATTAGGCGTTTCGC

2040 2050 2060 2070 2080 2090 2100

AAAACGATTAATCTGGCAGGGAAAACCCAAAAACACG**CAACTG**CCAGACGTATCCGAGATTATCAT**CAGG**
TTTTGCTAATTAGACCGTCCCTTTTGGGTTTTTGTGC**GTTGAC**GGTCTGCATAGGCTCTAATAGTA**GTCC**

2110 2120 2130 2140 2150 2160 2170

TGACCTATTGTTGGTCAGAAGAAAAGAAACAAACGGGCATGGATATGGATGGAGTGCGGAGAGAGGTTCC
ACTGGATAACAACCAGTCTTCTTTTCTTTGTTTGCCCGTACCTATACCTACCTCACGCCTCTCTCCAAGG

2180 2190 2200 2210 2220 2230 2240

CACTTTCAACCAGAACCGCCCACCAACCGACTACCTATTGTTGCTATTTTGACAGGATCGAGGTCCCAGG
GTGAAAGTTGGTCTTGGCGGGTGGTTGGCTGATGGATAACAACGATAAAACTGTCCTAGCTCCAGGGTCC

2250 2260 2270 2280 2290 2300 2310

TAGCGCAGAAAAATGCTAATTGGAAATTCCAGAACCTCTTTTCGAAGGCGTTGCAGATCTTTAGGTTATC
ATCGCGTCTTTTTTACGATTAACCTTTAAGGTCTTGGAGAAAAGCTTCCGCAACGTCTAGAAATCCAATAG

2320 2330 2340 2350 2360 2370 2380

CGCGAACCTTTCTCCCGAAAATATTGCAGCATAATGCCTAATCTGGGCCAGGATTTCGACTAAGCTGCGG
GCGCTTGGAAAGAGGGCTTTTATAACGTCGTATTACGGATTAGACCCGGTCCTAAAGCTGATTCGACGCC

2390 2400 2410 2420 2430 2440 2450

TCGCGTGGCACTCTTAACAAAGTGAACCGATCGAGCGATCGATCGATCCATCAGCTATCATCATCATCGT
AGCGCACCGTGAGAATTGTTTCACTTGGCTAGCTCGCTAGCTAGCTAGGTAGTCGATAGTAGTAGTAGCA

2460 2470 2480 2490 2500 2510 2520

CAAGATCAAACGTCAGGATCCTTTGTCGGCTCGGGACTTTGTTCCTTGACAATTTGGGCGAAGTAACTGG
GTTCTAGTTTGCAGTCCTAGGAAACAGCCGAGCCCTGAAACAAGGAACTGTTAAACCCGCTTCATTGACC

2530 2540 2550 2560 2570 2580 2590

GCCATCTTGCAGTGGCAT**CATTTG**TACTCTTTAGCCATGCTCGGCGCTTCTCGGCCAGTGTTTTC**CAATT**
CGGTAGAACGTCACCGTA**GTAAAC**ATGAGAAATCGGTACGAGCCGCGAAGAGCCGGTCACAAAAG**GTTAA**

2600 2610 2620 2630 2640 2650 2660

GTTTTGCCCAGTTTCCTTAGAAATTCAGCGAACTCCTCGAATTGATTTTCCCTCTAATGCTGCGAGTCTG
CAAAACGGGTCAAAGGAATCTTTAAGTCGCTTGAGGAGCTTAACTAAAAGGGAGATTACGACGCTCAGAC

2670 2680 2690 2700 2710 2720 2730

TGGAGTTAATTAGCCTTTCGATGCGGT**CACTTG**AAAGTGGGCTTGATCCCAAGATGACTGTGGAAACCAG
ACCTCAATTAATCGGAAAGCTACGCCA**GTGAAC**TTTCACCCGAACTAGGGTTCTACTGACACCTTTGGTC

2740 2750 2760 2770 2780 2790 2800

AGCATACGAATGAAAGTATATAATGTGAGGGTGCAACATATTGCACCAGAGTTTCGAAGGACCCATCGAA
TCGTATGCTTACTTTCATATATTACACTCCCACGTTGTATAACGTGGTCTCAAAGCTTCCTGGGTAGCTT

2810 2820 2830 2840 2850 2860 2870

TGGTTTCAGCATGTTAAAGTAGTGTAGTTTTATTATCGGAAAAGTAACTCTTCAGTAAAGGCGTTAATAT
ACCAAAGTCGTACAATTTCATCACATCAAAATAATAGCCTTTTCATTGAGAAGTCATTTCCGCAATTATA

2880 2890 2900 2910 2920 2930 2940

AGCACATTTCAAATCCCTTAATATGAAAATAATAATAAATAATAATAATGTGATACTTGCTTTGATTTCT
TCGTGTAAAGTTTAGGGAATTATACTTTTATTATTATTTATTATTATTACACTATGAACGAAACTAAAGA

2950 2960 2970 2980 2990 3000 3010

TCATGCTTTTTTGAATTTCGAACTTCTCGACGTCCTGTAATTTGGCCATTGAAAGCAGGAGAAAGCCACT
AGTACGAAAAAACTTAAAGCTTGAAGAGCTGCAGGACATTAAACCGGTAACTTTCGTCCTCTTTCGGTGA

3020 3030 3040 3050 3060 3070 3080

ATTAATACACCCACTTAAATTTGACCCAATAAAGTAAGTACAACCGCCTAAAATCGAAAGTCACAAACCG
TAATTATGTGGGTGAATTTAAACTGGGTTATTTCATTCATGTTGGCGGATTTTAGCTTTCAGTGTTTGGC

3090 3100 3110 3120 3130 3140 3150

CACAAGCA**CAATTG**TTGCTCTTTGTTCTGTCCGAAAAATGGGGTTTGCATCCAACCAGTCGAGCACTCCA
GTGTTCGT**GTTAAC**AACGAGAAACAAGACAGGCTTTTTACCCCAAACGTAGGTTGGTCAGCTCGTGAGGT

3160 3170 3180 3190 3200 3210 3220
CTCACTAGACCACAGACATGCAACCAAGATGATCCATGGCGATCTTCGATGGAGGCACACATTGTGGAAA
GAGTGATCTGGTGTCTGTACGTTGGTTCTACTAGGTACCGCTAGAAGCTACCTCCGTGTGTAACACCTTT

3230 3240 3250 3260 3270 3280 3290
ATATTCC**CATATG**ACAAGAGTCTTTTGTGCGTTCGCATTTTGACCTTGGCTAAATTTTGTTTGAATATAA
TATAAGG**GTATAC**TGTTCTCAGAAAACACGCAAGCGTAAAACTGGAACCGATTTAAAACAAACTTATATT

3300 3310 3320 3330 3340 3350 3360
AAAGTCAATTCCAGGGCAAGTAACAAAAGAGGTATGATTACAGACGGTAGC**CAAATG**ACAGGTCAAACAA
TTTCAGTTAAGGTCCCGTTCATTGTTTTCTCCATACTAATGTCTGCCATCG**GTTTAC**TGTCCAGTTTGTT

3370 3380 3390 3400 3410 3420 3430
AAGGAGTCGACCTGTAGCTGATACAGTAAATACCTGGGAGAAAGAGGGAGAGAGTTGCAGCCAGAGACAA
TTCCTCAGCTGGACATCGACTATGTCATTTATGGACCCTCTTTCTCCCTCTCTCAACGTCGGTCTCTGTT

3440 3450 3460 3470 3480 3490 3500
TTAAACGTCTTGACCGCAAGCCAATTTAGGGCACGGGCTGTGCTGCTGATCCTCAAGAGGTTGGCAATCG
AATTTGCAGAACTGGCGTTCGGTTAAATCCCGTGCCCGACACGACGACTAGGAGTTCTCCAACCGTTAGC

GG
CC

Primers Underlined, E-boxes in bold

Appendix D: E-boxes present in *amos* regulatory region

E-box	Conserved	Da contacts present	*E(spl)*
A1-CACAATTG G			
A2-GCCATATG T			
A3-TGCAGTT G			
A4- TCCACGTGT			Yes
A5-ATCAATTGC			
A6-AACACTTGT			
A7-TACACGTGA			Yes
A8- GACACTTGA			
A9- GGCAATTG	Yes		
B1-CAGAGATGT			
B2-GACAGTTCC			
B3-GCCAGTTCC			
B4-GTCACCTGA	Yes/Mutated	Yes	
B5-TACAAATGA	Yes		
B6-AACAATTGG	Yes		
C1-TTCAAGTGA	Yes/Mutated	Yes	
C2-AACAATTGT			
C3-GTCATATG			
C4-GTCATTTGG	Yes		

References

Allis, C. D., Waring, G. L., and Mahowald, A. P. (1977). Mass isolation of pole cells from Drosophila melanogaster. Dev Biol *56*, 372-381.

Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W., and White, K. P. (2002). Gene expression during the life cycle of Drosophila melanogaster. Science *297*, 2270-2275.

Atchley, W. R., and Fitch, W. M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci U S A *94*, 5172-5176.

Aubert, J., Stavridis, M. P., Tweedie, S., O'Reilly, M., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J. O., Roy, D., and Smith, A. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. Proc Natl Acad Sci U S A *100*, 11836-11841.

Bai, J., Chiu, W., Wang, J., Tzeng, T., Perrimon, N., and Hsu, J. (2001). The cell adhesion molecule Echinoid defines a new pathway that antagonizes the Drosophila EGF receptor signaling pathway. Development *128*, 591-601.

Bailey, A. M., and Posakony, J. W. (1995). Suppressor of hairless directly activates transcription of enhancer of split complex genes in response to Notch receptor activity. Genes Dev *9*, 2609-2622.

Barolo, S., Carver, L. A., and Posakony, J. W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. Biotechniques *29*, 726, 728, 730, 732.

Bellefroid, E. J., Kobbe, A., Gruss, P., Pieler, T., Gurdon, J. B., and Papalopulu, N. (1998). Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. Embo J *17*, 191-203.

Ben-Arie, N., Bellen, H. J., Armstrong, D. L., McCall, A. E., Gordadze, P. R., Guo, Q., Matzuk, M. M., and Zoghbi, H. Y. (1997). Math1 is essential for genesis of cerebellar granule neurons. Nature *390*, 169-172.

Ben-Arie, N., Hassan, B. A., Bermingham, N. A., Malicki, D. M., Armstrong, D., Matzuk, M., Bellen, H. J., and Zoghbi, H. Y. (2000). Functional conservation of atonal and Math1 in the CNS and PNS. Development *127*, 1039-1048.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49-59.

Bermingham, N. A., Hassan, B. A., Price, S. D., Vollrath, M. A., Ben-Arie, N., Eatock, R. A., Bellen, H. J., Lysakowski, A., and Zoghbi, H. Y. (1999). Math1: an essential gene for the generation of inner ear hair cells. Science *284*, 1837-1841.

Bermingham, N. A., Hassan, B. A., Wang, V. Y., Fernandez, M., Banfi, S., Bellen, H. J., Fritzsch, B., and Zoghbi, H. Y. (2001). Proprioceptor pathway development is dependent on Math1. Neuron *30*, 411-422.

Bhat, K. M. (1999). Segment polarity genes in neuroblast formation and identity specification during Drosophila neurogenesis. Bioessays *21*, 472-485.

Blader, P., Plessy, C., and Strahle, U. (2003). Multiple regulatory elements with spatially and temporally distinct activities control neurogenin1 expression in primary neurons of the zebrafish embryo. Mech Dev *120*, 211-218.

Blochlinger, K., Jan, L. Y., and Jan, Y. N. (1991). Transformation of sensory organ identity by ectopic expression of Cut in Drosophila. Genes Dev *5*, 1124-1135.

Bodmer, R., Barbel, S., Sheperd, S., Jack, J. W., Jan, L. Y., and Jan, Y. N. (1987). Transformation of sensory organs by mutations of the cut locus of D. melanogaster. Cell *51*, 293-307.

Brand, M., Jarman, A. P., Jan, L. Y., and Jan, Y. N. (1993). asense is a Drosophila neural precursor gene and is capable of initiating sense organ formation. Development *119*, 1-17.

Brown, N. L., Patel, S., Brzezinski, J., and Glaser, T. (2001). Math5 is required for retinal ganglion cell and optic nerve formation. Development *128*, 2497-2508.

Brunet, J. F., and Ghysen, A. (1999). Deconstructing cell determination: proneural genes and neuronal identity. Bioessays *21*, 313-318.

Bryant, Z., Subrahmanyan, L., Tworoger, M., LaTray, L., Liu, C. R., Li, M. J., van den Engh, G., and Ruohola-Baker, H. (1999). Characterization of differentially expressed genes in purified Drosophila follicle cells: toward a general strategy for cell type- specific developmental analysis. Proc Natl Acad Sci U S A *96*, 5559-5564.

Butler, M. J., Jacobsen, T. L., Cain, D. M., Jarman, M. G., Hubank, M., Whittle, J. R., Phillips, R., and Simcox, A. (2003). Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays. Development *130*, 659-670.

Cabrera, C. V., and Alonso, M. C. (1991). Transcriptional activation by heterodimers of the achaete-scute and daughterless gene products of Drosophila. Embo J *10*, 2965-2973.

Campuzano, S., Carramolino, L., Cabrera, C. V., Ruiz-Gomez, M., Villares, R., Boronat, A., and Modolell, J. (1985). Molecular genetics of the achaete-scute gene complex of D. melanogaster. Cell *40*, 327-338.

Casarosa, S., Fode, C., and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. Development *126*, 525-534.

Casso, D., Ramirez-Weber, F. A., and Kornberg, T. B. (1999). GFP-tagged balancer chromosomes for Drosophila melanogaster. Mech Dev *88*, 229-232.

Caudy, M., Vassin, H., Brand, M., Tuma, R., Jan, L. Y., and Jan, Y. N. (1988). daughterless, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. Cell *55*, 1061-1067.

Chan, Y. M., and Jan, Y. N. (1999). Conservation of neurogenic genes and mechanisms. Curr Opin Neurobiol *9*, 582-588.

Chien, C.-T., Hsiao, C.-D., Jan, L. Y., and Jan, Y. N. (1996). Neuronal type information encoded in the basic-helix-loop-helix domain of proneural genes. Proceedings of the National Academy of Sciences, USA *93*, 13239-13244.

Cronmiller, C., and Cummings, C. A. (1993). The daughterless gene product in Drosophila is a nuclear protein that is broadly expressed throughout the organism during development. Mech Dev *42*, 159-169.

Cubas, P., de Celis, J. F., Campuzano, S., and Modolell, J. (1991). Proneural clusters of achaete-scute expression and the generation of sensory organs in the Drosophila imaginal wing disc. Genes Dev *5*, 996-1008.

Culí, J., and Modolell, J. (1998). Proneural gene self-stimulation in neural precursors: an essential mechanism for sense organ development that is regulated by *Notch* signalling. Genes and Development *12*, 2036-2047.

Currie, D.A., Milner, M.J. and Evans C.W. (1988). The growth and differentiation in vitro of leg and wing imaginal disc cells from Drosophila melanogaster. Development *102, 805-814.*

Davis, R. L., and Weintraub, H. (1992). Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. Science *256*, 1027-1030.

de Celis, J. F., Barrio, R., and Kafatos, F. C. (1999). Regulation of the spalt/spalt-related gene complex and its function during sensory organ development in the Drosophila thorax. Development *126*, 2653-2662.

De Gregorio, E., Spellman, P. T., Rubin, G. M., and Lemaitre, B. (2001). Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc Natl Acad Sci U S A *98*, 12590-12595.

Dulac, C., and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. Cell *83*, 195-206.

Egger, B., Leemans, R., Loop, T., Kammermeier, L., Fan, Y., Radimerski, T., Strahm, M. C., Certa, U., and Reichert, H. (2002). Gliogenesis in Drosophila: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. Development *129*, 3295-3309.

Ellis, H. M., Spann, D. R., and Posakony, J. W. (1990). extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. Cell *61*, 27-38.

Ephrussi, A., Church, G. M., Tonegawa, S., and Gilbert, W. (1985). B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science *227*, 134-140.

Ernst, T., Hergenhahn, M., Kenzelmann, M., Cohen, C. D., Bonrouhi, M., Weninger, A., Klaren, R., Grone, E. F., Wiesel, M., Gudemann, C., *et al.* (2002). Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol *160*, 2169-2180.

Flores, G. V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., Noll, M., and Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. Cell *103*, 75-85.

Fode, C., Gradwohl, G., Morin, X., Dierich, A., LeMeur, M., Goridis, C., and Guillemot, F. (1998). The bHLH protein NEUROGENIN 2 is a determination factor for epibranchial placode-derived sensory neurons. Neuron *20*, 483-494.

Furlong, E. E., Andersen, E. C., Null, B., White, K. P., and Scott, M. P. (2001). Patterns of gene expression during Drosophila mesoderm development. Science *293*, 1629-1633.

Garcia-Garcia, M. J., Ramain, P., Simpson, P., and Modolell, J. (1999). Different contributions of pannier and wingless to the patterning of the dorsal mesothorax of Drosophila. Development *126*, 3523-3532.

Garrell, J., and Modolell, J. (1990). The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell *61*, 39-48.

Gomez-Skarmeta, J. L., Campuzano, S., and Modolell, J. (2003). Half a century of neural prepatterning: the story of a few bristles and many genes. Nat Rev Neurosci *4*, 587-598.

Gomez-Skarmeta, J. L., Diez del Corral, R., de la Calle-Mustienes, E., Ferre-Marco, D., and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell *85*, 95-105.

Gomez-Skarmeta, J. L., Glavic, A., de la Calle-Mustienes, E., Modolell, J., and Mayor, R. (1998). Xiro, a Xenopus homolog of the Drosophila Iroquois complex genes, controls development at the neural plate. Embo J *17*, 181-190.

Gomez-Skarmeta, J. L., and Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. Curr Opin Genet Dev *12*, 403-408.

Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., Ferres-Marco, D., Beamonte, D., and Modolell, J. (1995). Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. Genes Dev *9*, 1869-1882.

Goulding, S. E., White, N. M., and Jarman, A. P. (2000a). cato encodes a basic helix-loop-helix transcription factor implicated in the correct differentiation of Drosophila sense organs. Dev Biol *221*, 120-131.

Goulding, S. E., zur Lage, P., and Jarman, A. P. (2000b). amos, a proneural gene for Drosophila olfactory sense organs that is regulated by lozenge. Neuron *25*, 69-78.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R., and Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. Neuron *31*, 219-232.

Gradwohl, G., Fode, C., and Guillemot, F. (1996). Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. Dev Biol *180*, 227-241.

Guillemot, F., Lo, L. C., Johnson, J. E., Auerbach, A., Anderson, D. J., and Joyner, A. L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell *75*, 463-476.

Gupta, B. P., Flores, G. V., Banerjee, U., and Rodrigues, V. (1998). Patterning an epidermal field: Drosophila lozenge, a member of the AML-1/Runt family of transcription factors, specifies olfactory sense organ type in a dose-dependent manner. Dev Biol *203*, 400-411.

Gupta, B. P., and Rodrigues, V. (1995). Distinct mechanisms of action of the Lozenge locus in Drosophila eye and antennal development are suggested by the analysis of dominant enhancers. J Neurogenet *10*, 137-151.

Gupta, B. P., and Rodrigues, V. (1997). Atonal is a proneural gene for a subset of olfactory sense organs in Drosophila. Genes Cells *2*, 225-233.

Hassan, B. A., and Bellen, H. J. (2000). Doing the MATH: is the mouse a good model for fly development? Genes Dev *14*, 1852-1865.

Hassan, B. A., Bermingham, N. A., He, Y., Sun, Y., Jan, Y. N., Zoghbi, H. Y., and Bellen, H. J. (2000). atonal regulates neurite arborization but does not act as a proneural gene in the Drosophila brain. Neuron *25*, 549-561.

Helms, A. W., Abney, A. L., Ben-Arie, N., Zoghbi, H. Y., and Johnson, J. E. (2000). Autoregulation and multiple enhancers control Math1 expression in the developing nervous system. Development *127*, 1185-1196.

Huang, H. P., Liu, M., El-Hodiri, H. M., Chu, K., Jamrich, M., and Tsai, M. J. (2000a). Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol Cell Biol *20*, 3292-3307.

Huang, M. L., Hsu, C. H., and Chien, C. T. (2000b). The proneural gene amos promotes multiple dendritic neuron formation in the Drosophila peripheral nervous system. Neuron *25*, 57-67.

Hubank, M., and Schatz, D. G. (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. Nucleic Acids Res *22*, 5640-5648.

Hummel, T., Krukkert, K., Roos, J., Davis, G., and Klambt, C. (2000). Drosophila Futsch/22C10 is a MAP1B-like protein required for dendritic and axonal development. Neuron *26*, 357-370.

Ip, Y. T., Park, R. E., Kosman, D., Bier, E., and Levine, M. (1992). The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the Drosophila embryo. Genes Dev *6*, 1728-1739.

Jafar-Nejad, H., Acar, M., Nolo, R., Lacin, H., Pan, H., Parkhurst, S. M., and Bellen, H. J. (2003). Senseless acts as a binary switch during sensory organ precursor selection. Genes Dev *17*, 2966-2978.

Jarman, A. P., and Ahmed, I. (1998). The specificity of proneural genes in determining Drosophila sense organ identity. Mech Dev *76*, 117-125.

Jarman, A. P., Brand, M., Jan, L. Y., and Jan, Y. N. (1993a). The regulation and function of the helix-loop-helix gene, asense, in Drosophila neural precursors. Development *119*, 19-29.

Jarman, A. P., Grau, Y., Jan, L. Y., and Jan, Y. N. (1993b). atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system. Cell *73*, 1307-1321.

Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y., and Jan, Y. N. (1994). Atonal is the proneural gene for Drosophila photoreceptors. Nature *369*, 398-400.

Jarman, A. P., and Jan, Y. N. (1995). Multiple roles for proneural genes in *Drosophila* neurogenesis. In Neural Cell Specification: Molecular Mechanisms and Neurotherapeutic Implications, B. H. J. Juurlink, P. H. Krone, W. M. Kulyk, V. M. K. Verge, and J. R. Doucette, eds. (New York, Plenum Press), pp. 97-104.

Jarman, A. P., Sun, Y., Jan, L. Y., and Jan, Y. N. (1995). Role of the proneural gene, atonal, in formation of Drosophila chordotonal organs and photoreceptors. Development *121*, 2019-2030.

Jennings, B., De Celis, J., Delidakis, C., Preiss, A., and Bray, S. (1995). Role of Notch and achaete-scute complex in the expression of Enhancer of split bHLH proteins. Development *121*, 3745-3752.

Jhaveri, D., Sen, A., Reddy, G. V., and Rodrigues, V. (2000). Sense organ identity in the Drosophila antenna is specified by the expression of the proneural gene atonal. Mech Dev *99*, 101-111.

Jimenez, F., and Campos-Ortega, J. A. (1990). Defective neuroblast commitment in mutants of the achaete-scute complex and adjacent genes of D. melanogaster. Neuron *5*, 81-89.

Johnson, J. E., Birren, S. J., and Anderson, D. J. (1990). Two rat homologues of Drosophila achaete-scute specifically expressed in neuronal precursors. Nature *346*, 858-861.

Kanekar, S., Perron, M., Dorsky, R., Harris, W. A., Jan, L. Y., Jan, Y. N., and Vetter, M. L. (1997). Xath5 participates in a network of bHLH genes in the developing Xenopus retina. Neuron *19*, 981-994.

Kay, J. N., Finger-Baier, K. C., Roeser, T., Staub, W., and Baier, H. (2001). Retinal ganglion cell genesis requires lakritz, a Zebrafish atonal Homolog. Neuron *30*, 725-736.

Kim, P., Helms, A. W., Johnson, J. E., and Zimmerman, K. (1997). XATH-1, a vertebrate homolog of Drosophila atonal, induces a neuronal differentiation within ectodermal progenitors. Dev Biol *187*, 1-12.

Klebes, A., Biehs, B., Cifuentes, F., and Kornberg, T. B. (2002). Expression profiling of Drosophila imaginal discs. Genome Biol *3*, Research0038.

Kumar, J.P., Hsiung F., Powers M.A., and Moses K. (2003). Nuclear translocation of activated MAP kinase is developmentally regulated in the developing Drosophila eye. Development *130*. 3703-3714.

Kunisch, M., Haenlin, M., and Campos-Ortega, J. A. (1994). Lateral inhibition mediated by the Drosophila neurogenic gene delta is enhanced by proneural proteins. Proc Natl Acad Sci U S A *91*, 10139-10143.

Lai, E. C., Bodner, R., and Posakony, J. W. (2000). The enhancer of split complex of Drosophila includes four Notch-regulated members of the bearded gene family. Development *127*, 3441-3455.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell *66*, 305-315.

Ledent, V., and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res *11*, 754-770.

Lee, J. E. (1997). Basic helix-loop-helix genes in neural development. Curr Opin Neurobiol *7*, 13-20.

Leemans, R., Loop, T., Egger, B., He, H., Kammermeier, L., Hartmann, B., Certa, U., Reichert, H., and Hirth, F. (2001). Identification of candidate downstream genes for the homeodomain transcription factor Labial in Drosophila through oligonucleotide-array transcript imaging. Genome Biol *2*, RESEARCH0015.

Leyns, L., Danbly-Chaudiere, C., and Ghysen, A. (1989). Two different sets of *cis* elements regulate *scute* to establish two different sensory patterns. Wilhelm Roux's Arch Dev Biol *198*, 227-232.

Leyns, L., Gomez-Skarmeta, J. L., and Dambly-Chaudiere, C. (1996). iroquois: a prepattern gene that controls the formation of bristles on the thorax of Drosophila. Mech Dev *59*, 63-72.

Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction [see comments]. Science *257*, 967-971.

Luzzi, V., Mahadevappa, M., Raja, R., Warrington, J. A., and Watson, M. A. (2003). Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. J Mol Diagn *5*, 9-14.

Ma, Q., Chen, Z., del Barco Barrantes, I., de la Pompa, J. L., and Anderson, D. J. (1998). neurogenin1 is essential for the determination of neuronal precursors for proximal cranial sensory ganglia. Neuron *20*, 469-482.

Ma, Q., Fode, C., Guillemot, F., and Anderson, D. J. (1999). Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. Genes Dev *13*, 1717-1728.

Ma, Q., Kintner, C., and Anderson, D. J. (1996). Identification of neurogenin, a vertebrate neuronal determination gene. Cell *87*, 43-52.

Malatesta, P., Hartfuss, E., and Gotz, M. (2000). Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. Development *127*, 5253-5263.

Martinez, C., Modolell, J., and Garrell, J. (1993). Regulation of the proneural gene achaete by helix-loop-helix proteins. Mol Cell Biol *13*, 3514-3521.

Matter-Sadzinski, L., Matter, J. M., Ong, M. T., Hernandez, J., and Ballivet, M. (2001). Specification of neurotransmitter receptor identity in developing retina: the chick ATH5 promoter integrates the positive and negative effects of several bHLH proteins. Development *128*, 217-231.

McNeill, H., Yang, C. H., Brodsky, M., Ungos, J., and Simon, M. A. (1997). mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. Genes Dev *11*, 1073-1082.

Mlodzik, M., Baker, N. E., and Rubin, G. M. (1990). Isolation and expression of scabrous, a gene regulating neurogenesis in Drosophila. Genes Dev *4*, 1848-1861.

Molkentin, J. D., and Olson, E. N. (1996). Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. Proc Natl Acad Sci U S A *93*, 9366-9373.

Molkentin, J.D., and Olson, E.N. (1996). Defining the regulatory networks for muscle development. Curr Opin in Gene and Dev 1996.

Moore, A. W., Barbel, S., Jan, L. Y., and Jan, Y. N. (2000). A genomewide survey of basic helix-loop-helix factors in Drosophila. Proc Natl Acad Sci U S A *97*, 10436-10441.

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., and et al. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell *58*, 537-544.

Neufeld, T. P., de la Cruz, A. F., Johnston, L. A., and Edgar, B. A. (1998). Coordination of growth and cell division in the Drosophila wing. Cell *93*, 1183-1193.

Nolo, R., Abbott, L. A., and Bellen, H. J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. Cell *102*, 349-362.

Oellers, N., Dehio, M., and Knust, E. (1994). bHLH proteins encoded by the Enhancer of split complex of Drosophila negatively interfere with transcriptional activation mediated by proneural genes. Mol Gen Genet *244*, 465-473.

Ohyama, H., Zhang, X., Kohno, Y., Alevizos, I., Posner, M., Wong, D. T., and Todd, R. (2000). Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. Biotechniques *29*, 530-536.

Okabe, M., and Okano, H. (1997). Two-step induction of chordotonal organ precursors in Drosophila embryogenesis. Development *124*, 1045-1053.

Pi, H., Wu, H. J., and Chien, C. T. (2001). A dual function of phyllopod in Drosophila external sensory organ development: cell fate specification of sensory organ precursor and its progeny. Development *128*, 2699-2710.

Porcher, C., Liao, E. C., Fujiwara, Y., Zon, L. I., and Orkin, S. H. (1999). Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. Development *126*, 4603-4615.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F. W., and Orkin, S. H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47-57.

Powell, L. M., zur Lage, P., Prentice, D. R., Senthinathan, B., and Jarman, A. P. The proneural proteins Atonal and Scute can regulate neural target genes through different E-box binding sites. Paper submitted.

Prober, D. A., and Edgar, B. A. (2000). Ras1 promotes cellular growth in the Drosophila wing. Cell *100*, 435-446.

Prober, D. A., and Edgar, B. A. (2002). Interactions between Ras1, dMyc, and dPI3K signaling in the developing Drosophila wing. Genes Dev *16*, 2286-2299.

Ramain, P., Khechumian, R., Khechumian, K., Arbogast, N., Ackermann, C., and Heitzler, P. (2000). Interactions between chip and the achaete/scute-daughterless heterodimers are required for pannier-driven proneural patterning. Mol Cell *6*, 781-790.

Ray, K., and Rodrigues, V. (1995). Cellular events during development of the olfactory sense organs in *Drosophila melanogaster*. Developmental Biology *167*, 426-438.

Reddy, G. V., Gupta, B., Ray, K., and Rodrigues, V. (1997). Development of the Drosophila olfactory sense organs utilizes cell-cell interactions as well as lineage. Development *124*, 703-712.

Rodriguez, I., Hernandez, R., Modolell, J., and Ruiz-Gomez, M. (1990). Competence to develop sensory organs is temporally and spatially regulated in Drosophila epidermal primordia. Embo J *9*, 3583-3592.

Romani, S., Campuzano, S., Macagno, E. R., and Modolell, J. (1989). Expression of achaete and scute genes in Drosophila imaginal discs and their function in sensory organ development. Genes Dev *3*, 997-1007.

Rosay, P., Colas, J. F., and Maroteaux, L. (1995). Dual organisation of the Drosophila neuropeptide receptor NKD gene promoter. Mechanisms of Development *51*, 329-339.

Ruiz-Gomez, M., and Ghysen, A. (1993). The expression and role of a proneural gene, achaete, in the development of the larval nervous system of Drosophila. Embo J *12*, 1121-1130.

Ruiz-Gomez, M., and Modolell, J. (1987). Deletion analysis of the achaete-scute locus of Drosophila melanogaster. Genes Dev *1*, 1238-1246.

Sato, M., Kojima, T., Michiue, T., and Saigo, K. (1999). Bar homeobox genes are latitudinal prepattern genes in the developing Drosophila notum whose expression is regulated by the concerted functions of decapentaplegic and wingless. Development *126*, 1457-1466.

Scardigli, R., Baumer, N., Gruss, P., Guillemot, F., and Le Roux, I. (2003). Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. Development *130*, 3269-3281.

Scardigli, R., Schuurmans, C., Gradwohl, G., and Guillemot, F. (2001). Crossregulation between Neurogenin2 and pathways specifying neuronal identity in the spinal cord. Neuron *31*, 203-217.

Seale, P., Sabourin, L. A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., and Rudnicki, M. A. (2000). Pax7 is required for the specification of myogenic satellite cells. Cell *102*, 777-786.

Sen, A., Reddy, G. V., and Rodrigues, V. (2003). Combinatorial expression of Prospero, Seven-up, and Elav identifies progenitor cell types during sense-organ differentiation in the Drosophila antenna. Dev Biol *254*, 79-92.

Sharma, Y., Cheung, U, Larsen, E. W., Eberl D. (2002). PPTGal4, a convenient Gal4 P-element vector for testing expression of enhancer fragments in *Drosophila*. Genesis 34,1-1, 115-8.

Simpson, P. (1997). Notch signalling in development: on equivalence groups and asymmetric developmental potential. CURRENT OPINION IN GENETICS & DEVELOPMENT *7*, 537-542.

Singson, A., Leviten, M. W., Bang, A. G., Hua, X. H., and Posakony, J. W. (1994). Direct downstream targets of proneural activators in the imaginal disc include genes involved in lateral inhibitory signaling. Genes and Development *8*, 2058-2071.

Skeath, J. B., and Carroll, S. B. (1991). Regulation of achaete-scute gene expression and sensory organ pattern formation in the Drosophila wing. Genes Dev *5*, 984-995.

Skeath, J. B., and Carroll, S. B. (1992). Regulation of proneural gene expression and cell fate during neuroblast segregation in the Drosophila embryo. Development *114*, 939-946.

Skeath, J. B., Panganiban, G., Selegue, J., and Carroll, S. B. (1992). Gene regulation in two dimensions: the proneural achaete and scute genes are controlled by combinations of axis-patterning genes through a common intergenic control region. Genes Dev *6*, 2606-2619.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., and Levine, M. (2002). Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. Cell *111*, 687-701.

Stocker, R. F., Gendre, N., and Batterham, P. (1993). Analysis of the antennal phenotype in the Drosophila mutant lozenge. J Neurogenet *9*, 29-53.

Sun, Y., Jan, L. Y., and Jan, Y. N. (1998). Transcriptional regulation of *atonal* during development of the *Drosophila* peripheral nervous system. Development *125*, 3731-3740.

Turner, D. L., and Weintraub, H. (1994). Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate. Genes Dev *8*, 1434-1447.

Van Doren, M., Ellis, H. M., and Posakony, J. W. (1991). The Drosophila extramacrochaetae protein antagonizes sequence-specific DNA binding by daughterless/achaete-scute protein complexes. Development *113*, 245-255.

Van Doren, M., Powell, P. A., Pasternak, D., Singson, A., and Posakony, J. W. (1992). Spatial regulation of proneural gene activity: auto- and cross-activation of achaete is antagonized by extramacrochaetae. Genes Dev *6*, 2592-2605.

Wang, S. W., Kim, B. S., Ding, K., Wang, H., Sun, D., Johnson, R. L., Klein, W. H., and Gan, L. (2001). Requirement for math5 in the development of retinal ganglion cells. Genes Dev *15*, 24-29.

White, K. P., Rifkin, S. A., Hurban, P., and Hogness, D. S. (1999). Microarray analysis of Drosophila development during metamorphosis. Science *286*, 2179-2184.

Wyss, C. (1982). Ecdysterone, insulin and fly extract needed for the proliferation of normal Drosophila cells in defined medium. Exp Cell Res *139*, 297-307.

Xu, C., Kauffmann, R. C., Zhang, J., Kladny, S., and Carthew, R. W. (2000). Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. Cell *103*, 87-97.

Zipursky, S. L., Venkatesh, T. R., Teplow, D. B., and Benzer, S. (1984). Neuronal development in the Drosophila retina: monoclonal antibodies as molecular probes. Cell *36*, 15-26.

zur Lage, P., Jan, Y. N., and Jarman, A. P. (1997). Requirement for EGF receptor signalling in neural recruitment during formation of *Drosophila* chordotonal sense organ clusters. Current Biology *7*, 166-175.

zur Lage, P., and Jarman, A. P. (1999). Antagonism of EGFR and notch signalling in the reiterative recruitment of Drosophila adult chordotonal sense organ precursors. Development *126*, 3149-3157.

zur Lage, P. I., Prentice, D. R., Holohan, E. E., and Jarman, A. P. (2003). The Drosophila proneural gene amos promotes olfactory sensillum formation and suppresses bristle formation. Development *130*, 4683-4693.

The *Drosophila* proneural gene *amos* promotes olfactory sensillum formation and suppresses bristle formation

Petra I. zur Lage, David R. A. Prentice, Eimear E. Holohan and Andrew P. Jarman*

The Wellcome Trust Centre for Cell Biology, Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

*Author for correspondence (e-mail: andrew.jarman@ed.ac.uk)

Accepted 20 June 2003

Development 130, 4683-4693

doi:10.1242/dev.00680

Summary

Proneural genes encode basic-helix-loop-helix (bHLH) transcription factors required for neural precursor specification. Recently *amos* was identified as a new candidate *Drosophila* proneural gene related to *atonal*. Having isolated the first specific *amos* loss-of-function mutations, we show definitively that *amos* is required to specify the precursors of two classes of olfactory sensilla. Unlike other known proneural mutations, a novel characteristic of *amos* loss of function is the appearance of ectopic sensory bristles in addition to loss of olfactory sensilla, owing to the inappropriate function of *scute*. This supports a model of inhibitory interactions between proneural genes, whereby *ato*-like genes (*amos* and *ato*) must suppress sensory bristle fate as well as promote alternative sense organ subtypes.

Key words: Proneural, bHLH, *Drosophila*, *amos*, Neurogenesis, Gene regulation

Introduction

The sequence and structure of the bHLH domain is highly conserved, and yet transcription factors of this family play a variety of roles in neurogenesis in a range of organisms (Bertrand et al., 2002). These roles include conferring neuronal competence, directing neural precursor specification, directing neuronal subtype specification and triggering neuronal differentiation. Dissecting bHLH gene functions and interactions is an important and challenging task, and the *Drosophila* PNS provides a good model in which to do this. Here, proneural bHLH genes are required for sense organ precursor (SOP) specification (Hassan and Bellen, 2000). These genes include *achaete* (*ac*) and *scute* (*sc*), from the *Achaete-scute Complex* (*ASC*), and *atonal* (*ato*), as well as the candidate proneural gene *amos* (Hassan and Bellen, 2000). These proneural proteins seem to combine two functions: promoting SOP specification, and providing these SOPs with information concerning neuronal subtype (Jarman and Ahmed, 1998). It is thought that vertebrate proneural gene homologues also have functions in neural progenitor specification and neural subtype identity (Hassan and Bellen, 2000; Bertrand et al., 2002).

bHLH functions depend on both intrinsic protein properties and extrinsic factors (Bertrand et al., 2002). Comparisons of protein capabilities, particularly by assaying the effect of misexpression on neural development, have shown evidence for intrinsic differences between closely related bHLH proteins, suggesting that they regulate distinct target genes (Jarman and Ahmed, 1998). However, bHLH protein specificity is also very dependent on extrinsic modifying factors. Ato has been well characterised and illustrates well the complexity of defining the intrinsic specificity of proneural proteins. In most of the developing ectoderm, Ato is required for chordotonal (stretch receptor) SOP specification. Ectopic expression of *ato* leads to ectopic chordotonal SOP formation (Jarman et al., 1993). In this property, it differs from Ac and Sc, which are necessary and sufficient for external sense organ (bristle) SOPs. This points to intrinsic differences in protein properties. However, the function of Ato is clearly also very context dependent. In addition to specifying chordotonal organs, Ato is also required for R8 photoreceptors in the eye (Jarman et al., 1994), and for one subset of olfactory sensilla (sensilla coeloconica) in the antenna (Gupta and Rodrigues, 1997). Moreover, in a group of CNS neurons, Ato regulates neurite arborization (Hassan et al., 2000). It is not known how the response to Ato is modified in these different regions.

We have argued for a specific mechanism by which proneural proteins specify neural subtype: SOPs may be biased to become external sense organs and, consequently, Ac/Sc promotes a default neural fate, whereas Ato must actively impose alternative neural fates (Jarman and Ahmed, 1998). This idea is based on two apparently paradoxical outcomes of misexpression experiments. Under certain very defined conditions, *ato* misexpression can transform *existing* bristle SOPs to chordotonal organs, thereby revealing an intrinsic ability of Ato (Jarman and Ahmed, 1998). However, in most contexts, *ato* misexpression induces a mixture of ectopic chordotonal and bristle SOPs (Jarman et al., 1993), suggesting that in many circumstances Ato can specify SOPs but may often fail to provide subtype information. This suggests that the two proneural roles are separable in misexpression studies, and it also gives the appearance that Ato function is more sensitive to cell context than is Ac/Sc function. Similar controlled misexpression data for vertebrate bHLH genes have recently been reported, which support an entirely analogous situation in which *neurogenin* (*ato* homologue) is more context

sensitive than *Mash1* (*ac/sc* homologue) (Lo et al., 2002) (see also Parras et al., 2002). But teasing out these functions is complicated and misexpression data could be misleading. There is no corroborative evidence from loss-of-function mutations in *Drosophila* as known proneural mutations always cause loss of SOP subsets, and so questions concerning the neural identity of SOPs are hard to approach through loss-of-function studies.

Recently, we and others described a new candidate proneural gene, *amos* (Goulding et al., 2000; Huang et al., 2000). Amos protein possesses a very similar bHLH domain to that of Ato, suggesting there may be functional similarities with Ato that set this gene pair apart from *ac/sc*. We provided strong but indirect evidence that *amos* is the proneural gene for the *ato*-independent classes of olfactory sensillum (sensilla basiconica and trichodea) (Goulding et al., 2000). Here, we report a detailed analysis of *amos* expression and function, including the first isolation and characterisation of specific *amos* mutations. We find that Amos protein is expressed in, and is required for, a late wave of olfactory SOPs in the antenna. These are the precursors for sensilla basiconica and trichodea, proving that *amos* is the proneural gene for these subtypes. However, an unexpected aspect of the mutant phenotype was the appearance of ectopic sensory bristles in place of the olfactory sensilla on the antenna. This replacement of sense organs rather than complete absence is unprecedented for a *Drosophila* proneural gene mutation. Our analysis suggests that loss of *amos* results in loss of olfactory sensilla and concomitant derepression of *ac/sc* leading to formation of external sense organ SOPs. This phenotype supports the argument that the *ato*-like proneural genes (*amos* and *ato*) suppress external sense organ fate as well as promote alternative neural fates.

Materials and methods

Fly stocks

Wild-type flies are generally Oregon R and pr^1, as appropriate. UAS-*amos* is described by Goulding et al. (Goulding et al., 2000). sc^{10-1}, ase^1 and lz^{34} are described by Lindsley and Zimm (Lindsley and Zimm, 1992). Deficiencies and mutants were obtained from the Umea stock centre. For *ato/amos* double-mutant analysis, $amos^1$ clones were induced in an ato^1 background by the FLP/FRT method using *eyelessFLP* (Newsome et al., 2000). The flies had the following genotype: *y w eyFLP; amos*^{1} *pr*^{1} *FRT40A/2×nlsGFP FRT40A; ato*^{1}. Clones were recognised by their sensillum phenotype.

Mutagenesis

$amos^1$ was isolated in an F2 screen for mutations that failed to complement a deficiency of the *amos* region [*Df(2L)M36F-S6* (Goulding et al., 2000)]. pr^1 male flies were mutagenised with 25-30 mM EMS. Mutagenised lines were collected over a CyO balancer and individually tested for complementation with *Df(2L)M36F-S6*. 4500 mutagenised lines were screened. $amos^2$ and $amos^3$ were isolated in a subsequent F1 screen of 25,000 flies using $amos^1$. PCR isolation of the ORFs and sequencing were by standard techniques.

Amos enhancer construct

A 3.6 kb fragment upstream of the *amos* start site was amplified by PCR and cloned into the transformation vector pTLGal4 (a gift of B. Hassan). Transformant flies were made by microinjection into syncytial blastoderm embryos. These were crossed to UAS-*GFP* or UAS-*nlsGFP* lines for assessment of enhancer activity.

Immunohistochemistry

Antibody staining of pupal antennae was carried out as previously described (Goulding et al., 2000). Pupae were staged by collecting at the time of puparium formation and then ageing on moist filter paper at 25°C before dissection. Antibodies used were: Cut (1:100), Ac (1:50), 22C10 (1:200) and Elav (1:200) (all from the Developmental Biology Hybridoma Bank, Iowa); Sens (1:6250) (Nolo et al., 2000); and Pros (1:200). Anti-Amos antibodies were raised in rabbits, using full-length His_6-tagged Amos protein expressed in *E. coli*, and purified by adsorption to nickel-agarose under denaturing conditions. Anti-Amos antibodies were used at 1:1250 after pre-adsorption against wild-type embryos. RNA in situ hybridisation was done according to standard protocols using digoxigenin-labelled *sc* cDNA. RNA/protein double labellings were carried out by initially detecting RNA using anti-digoxigenin-POD and an Alexa Fluor 488 tyramide substrate (Molecular Probes), followed by antibody staining. Microscopy analysis was carried out using an Olympus AX70 or Leica LCS-SP system.

Results

amos mutations result in loss of olfactory sensilla and the appearance of mechanosensory bristles

We generated three mutant alleles of *amos* in an EMS screen (Table 1). $amos^1$ is predicted to result in a protein truncation that removes the second half of bHLH helix 2 and the C-terminal region thereafter. $amos^2$ is a missense mutation that changes a Ser to an Asn in helix 1 of the bHLH domain. This position is not part of the bHLH core consensus sequence and is not predicted to affect directly DNA binding or dimerisation. Moreover, Asn is found in this position in the *ato* bHLH domain, and so the effect of this mutation would be predicted to be mild. $amos^3$ contains a 230 bp deletion within the ORF, which also causes a frame-shift that brings a spurious downstream stop codon in frame. This allele gives a predicted peptide of 74 amino acids, of which only the first 30 are shared with *amos*. It therefore lacks the entire bHLH domain and is likely to be a null.

Consistent with previous RNAi experiments (Huang et al., 2000), $amos^1$ mutant embryos lack two dorsal sensory neurons per segment, including the dorsal bipolar dendritic neuron (D.R.A.P. and A.P.J., unpublished). Nevertheless, all *amos* alleles are adult viable as homozygotes and hemizygotes. The antennae of mutant adult flies were mounted and examined by light microscopy in order to quantify the number and type of olfactory sensilla. Compared with wild-type (Fig. 1A and Table 2) (Carlson, 1996), *amos* mutant antennae carried dramatically reduced numbers of sensilla and, as a consequence, the third segment is significantly smaller (Fig. 1C,D). In particular,

Table 1. Molecular basis of *amos* mutations

Allele	DNA	Protein	Predicted effect	Phenotype
$amos^1$	C_{550}>T_{550}	Q_{184}>Stop	Truncates the bHLH domain	Strong hypomorph/null
$amos^2$	G_{458}>A_{458}	S_{153}>N_{153}	Substitution in bHLH domain	Moderate hypomorph
$amos^3$	230 bp deletion+frameshift		Severely truncated protein	Null?

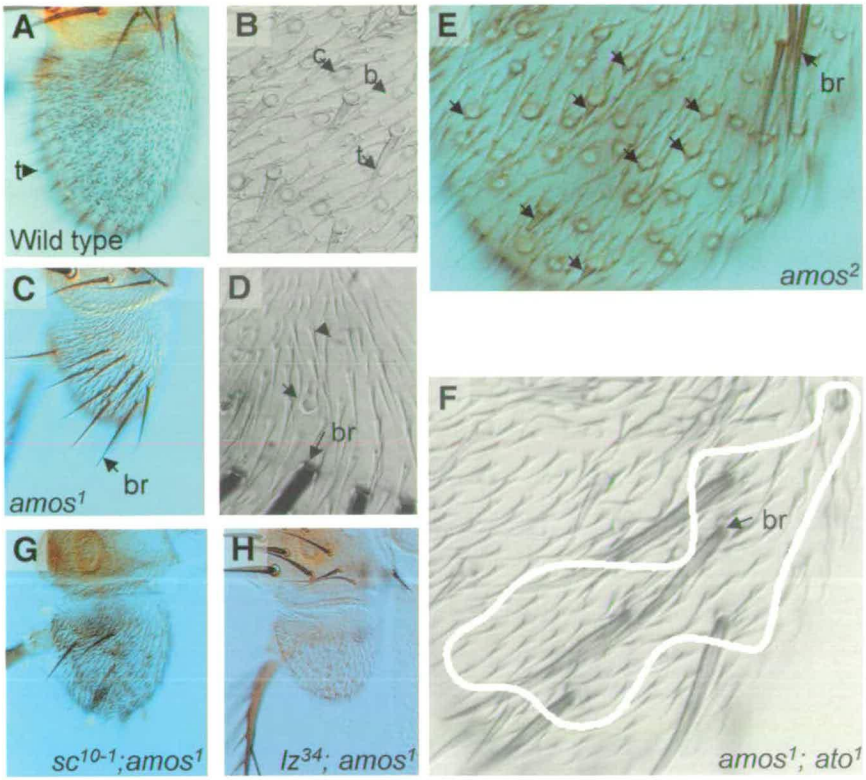


Fig. 1. Antennal defects in *amos* mutants. Third antennal segments are shown. (A) Wild-type antenna, with trichodea indicated (t). (B) Higher power view with examples marked of sensilla coeloconica (c), basiconica (b) and trichodea (t). (C) *amos*^{1}*/Df(2L)M36-S6*. Third segment is reduced because of missing basiconica and trichodea. Ectopic mechanosensory bristles are indicated (br). (D) Higher power view showing an abnormal domed sensillum (arrow) and a normal sensillum coeloconicum (arrowhead). Sensilla are very sparse. (E) *amos*^{2} with numerous abnormal sensilla (arrows). (F) Double mutant for *amos* and *ato* (clone of *amos*^{1} tissue in *ato*^{1} fly: *eyFLP; FRT-amos*^{1}*/FRT-nlsGFP; ato*^{1}). The clone patch contains no sense organs except bristles. (G) Double mutant for *amos* and *sc*$^{10\text{-}1}$ (which removes both *ac* and *sc* function). The extra bristles of *amos*^{1} are largely dependent on *ac/sc* function. (H) Double mutant for *amos* and *lz*. The extra bristles of *amos*^{1} are absent and therefore depend on *lz* function.

sensilla basiconica and trichodea were completely absent in the probable genetic nulls, *amos*^{1} and *amos*^{3}, whereas sensilla coeloconica appeared unaffected (Table 2). These phenotypes support the assertion that *amos* is the proneural gene for sensilla basiconica and trichodea, whereas *ato* is the proneural gene for sensilla coeloconica. However, mutant antennae exhibit a further unexpected phenotype. In wild type, the third segment bears only olfactory sensilla; in *amos* mutant flies, this segment bears a number of ectopic external sensory bristles and other abnormally structured sensilla (Fig. 1C,D and Table 2). These bristles do not have bracts (unlike bristles on the leg), and so this phenotype does not represent a transformation of antenna to leg (c.f. Johnston et al., 1998). These phenotypes are highly unusual as a characteristic of all other loss-of-function proneural gene mutations is that they cause the loss of sense organ subsets without the concomitant appearance of new or abnormal sensory structures. Therefore, the *amos* null phenotype is unique for a *Drosophila* proneural gene.

Given its subtle molecular basis, the putative hypomorph *amos*^{2} has a surprisingly strong phenotype: it has no sensilla trichodea, and sensilla basiconica are reduced very substantially (Table 2). There are also fewer ectopic bristles than in the null alleles, but there are many sensilla of unusual morphology. In the case of this allele, these seem to represent intermediates between sensilla basiconica/trichodea and external sense organs (Fig. 1E).

Late pupal antennae were stained with a sensory neuron marker, MAb22C10, to visualise olfactory receptor neurons (ORNs). Olfactory sensilla are innervated by multiple sensory neurons (Shanbhag et al., 1999), which can be seen as groups in the wild-type antenna (Fig. 2A). *amos* mutant antennae have many fewer neuronal groups, corresponding in number to the sensilla coeloconica and the bristles (Fig. 2B). There are instances of sensilla innervated by a single neuron, which appear to correspond to the ectopic bristles (Fig. 2C,D). In wild-type flies, ORN axons form three olfactory nerves leading to the antennal lobe of the brain (Jhaveri et al., 2000b) (Fig. 2E). In *amos* mutant antennae, all three antennal nerves are still

Table 2. Sensillum numbers on adult antennae

Genotype	Basiconica	Trichodea	Coeloconica	Mechanosensory bristle	Mixed*
Wild type	177.5±8.6	114.5±3.5	70.3±1.5	0	0
amos^{1}	0	0	63.8±9.3	14.4±4.2	7.0±2.1
amos^{1}*/Df(2L)M36F-S6*	0	0	53.5±5.0	14.5±3.0	2.2±2.2
amos^{2}	9.6±5.6	0	62.8±7.1	13.3±4.5	15.2±3.8
amos^{2}*/Df(2L)M36F-S6*	5.8±3.4	0	52.0±1.4	12.5±3.9	21.5±2.6
amos^{3}*/Df(2L)M36F-S6*	0	0	72.6±6.4	20.8±0.8	3.6±2.2
lz^{34}; *amos*^{1}*/Df(2L)M36F-S6*	0	0	56.7±3.8	0	0
sc$^{10\text{-}1}$; *amos*^{1}	0	0	83.5±0.7	6.0±1.4	10.5±2.2
ase^{1}; *amos*^{1}*/Df(2L)M36F-S6*	0	0	58.2±8.3	14.8±2.2	6.8±2.9

Values are sensilla per antenna±s.d. (number of antenna scored, n=4-11).
'Mixed' refers to sensilla of undefinable morphology (*amos*^{1} and *amos*^{3}) or intermediate olfactory/bristle morphology (*amos*^{2}).

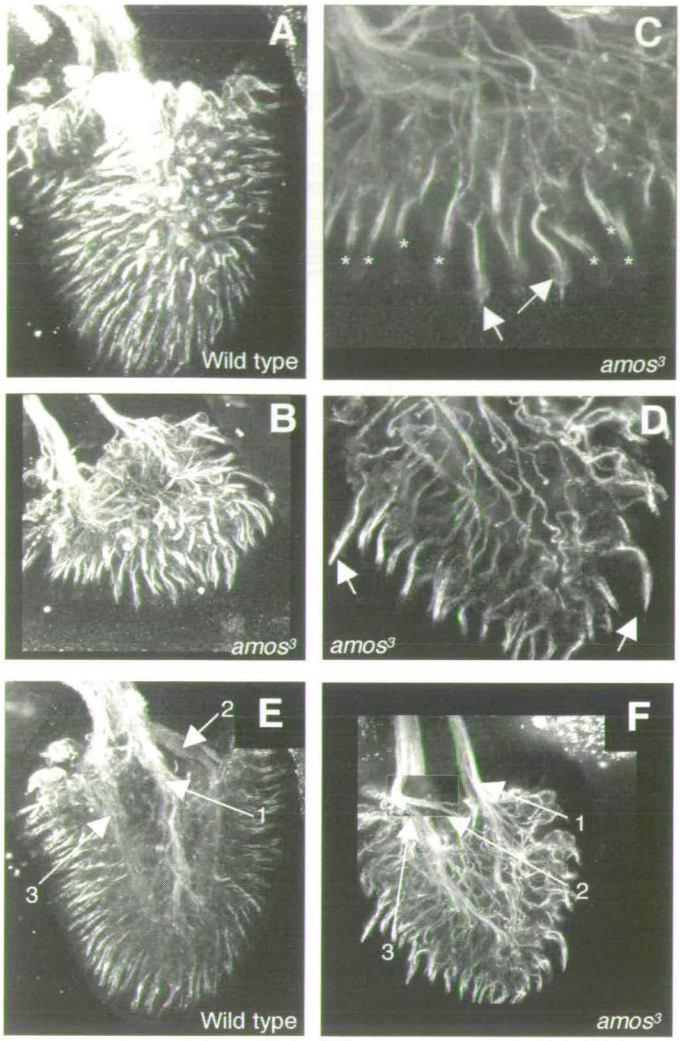


Fig. 2. Olfactory receptor neurons in *amos* mutants. Confocal projection images of late pupal antennae stained to detect sensory neurons. (A) Wild type. Clusters of cell bodies and their dendrites can be seen. (B) *amos*^{3} mutant showing far fewer clusters. (C,D) Higher magnification views. Although most ORNs are clustered, as seen by the multiple dendrites (*) (representing sensilla coeloconica), some sensilla appear to be mono-innervated (arrows) and may represent the bristles. (E) Wild-type confocal section showing the three olfactory nerve bundles. (F) Confocal section of *amos* mutant, with the three bundles labelled (a small section has been pasted in from another confocal plane to show clearly the second bundle).

present, although consisting of fewer axons as expected (comprising the axons of *ato*-dependent ORNs) (Fig. 2F). Although thinner, the fascicles appear normal in structure and location. Thus, in contrast to *ato* (Jhaveri et al., 2000b), mutations of *amos* do not cause defects in routing or fasciculation of the olfactory nerves. This supports the conclusion that the *ato*-dependent sensory lineage provides the information for fasciculation of these nerves.

amos expression prefigures a late, *ato*-independent subset of olfactory precursors

Olfactory precursors arise in the pupal antennal imaginal disc over an extended period of time (Ray and Rodrigues, 1995). Given their high density, the appearance of olfactory precursors is complex and incompletely characterised. We initially characterised the evolution of this pattern by studying Senseless (Sens; Lyra – FlyBase) expression, which is a faithful indicator of proneural-derived sensory precursors and is probably a direct target of proneural proteins (Nolo et al., 2000). We found that precursor formation occurs in three waves. First, Sens expression begins a few hours before puparium formation (BPF) in an outer semicircle of cells (Fig. 3A). A second wave begins at 0-4 hours after puparium formation (APF) to give a very characteristic pattern, including three semicircles of precursors (Fig. 3B). After this, a third wave appears over an extended period of time, with increasing numbers of cells appearing intercalated between the early precursors until no spatial pattern features can be observed (Fig. 3C,D).

Using a polyclonal antibody raised against the entire Amos protein, we determined that *amos* expression begins at puparium formation in three distinct semicircles and then continues for the next 16 hours, with the semicircles becoming indistinct by around 8 hours APF (Fig. 3E-H). The characteristic early waves of SOPs arise between the Amos domains of expression and do not show overlap with Amos expression (Fig. 3I-L). However, the third wave of SOPs appears to arise from the Amos expression domains. These late SOPs co-express Amos, and their nuclei lie beneath the Amos expression domains, consistent with these cells being olfactory SOPs (Fig. 4E,F and data not shown). Unusually, overlying *amos* proneural cluster expression is evidently not affected by lateral inhibition upon the appearance of the SOPs. *ato* is expressed much earlier than *amos*. All wave 1 and 2 SOPs appear to express Ato or to have arisen from Ato-expressing cells (see also Jhaveri et al., 2000a) (Fig. 4A-D). Consistent with this, the entire early SOP pattern is missing in antennal discs from *ato*^{1} mutant pupae (Fig. 5A-C). SOPs only begin to appear between 4 and 8 hours APF, corresponding to the third wave of precursors. These coincide very precisely with Amos expression, which itself appears unaffected (Fig. 5D-F).

In summary, there are three waves of olfactory precursor formation (Fig. 4G). The first and second waves are well defined, giving rise to the sensilla coeloconica of the sacculus and the antennal surface, respectively. These precursors express and require *ato*. The third wave of precursors is much more extensive and has little obvious pattern, giving rise to the more numerous sensilla basiconica and trichodea. Amos is expressed in a pattern entirely consistent with it being the proneural gene for the late wave precursors. Expression of Amos is complementary with that of Ato and is independent of Ato function. Thus, Ato- and Amos-expressing SOPs show a degree of spatial and temporal separation.

amos appears to be expressed in proneural domains and then in SOPs. For *sc* and *ato*, these two phases of expression are driven by separate enhancers, and SOP-specific enhancers have been identified (Culí and Modolell, 1998; Sun et al., 1998). A 3.6 kb fragment upstream from *amos* was found to support GFP reporter gene expression in the pupal third antennal segment. Comparison with Amos and Sens expression showed that GFP coincides with the Amos but not Ato SOPs (Fig. 6A,B). This fragment therefore contains an *amos* SOP enhancer. Perduring GFP expression driven by the enhancer

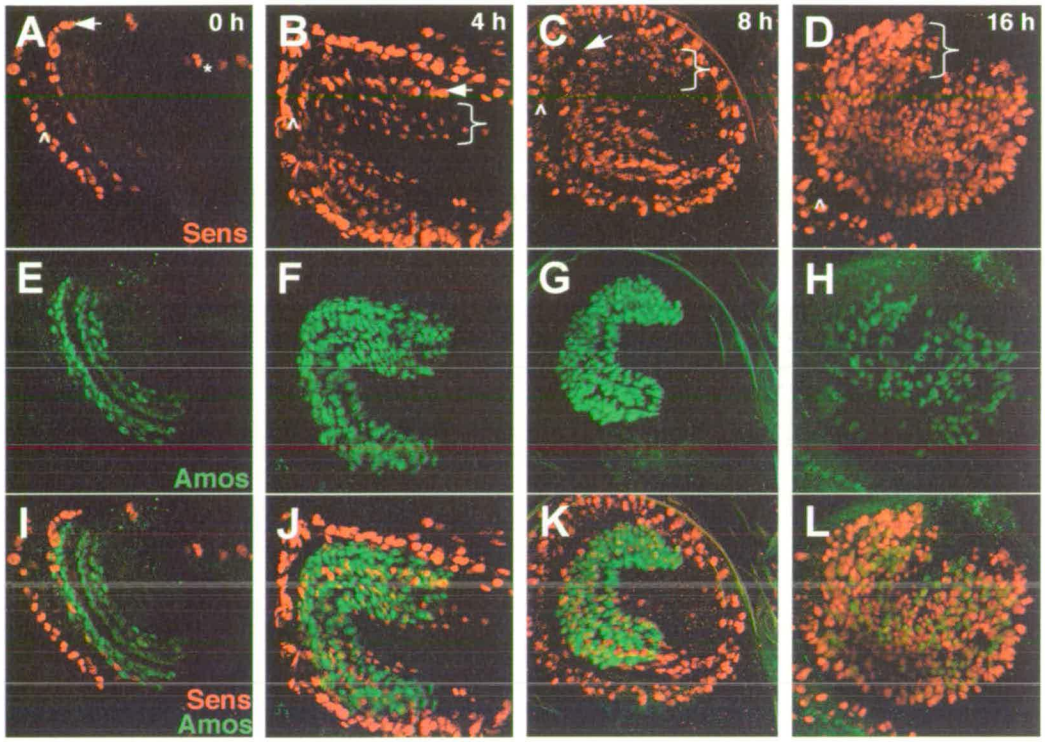


Fig. 3. Amos expression during olfactory precursor formation. (A-L) Time course of Amos protein expression relative to olfactory precursor formation. In all cases we concentrate on precursors in the third segment, although a large number of chordotonal precursors are also visible in the surrounding second segment (carets; see also the summary in Fig. 4). (A,E,I) At 0 hours APF, the first wave of precursors appear (arrow). (B,F,J) At 4 hours APF, the second wave of precursors appears in a highly characteristic pattern (bracket). (C,G,K) At 8 hours APF, the third wave of precursors accumulate between the rows of the second wave, eventually obscuring any clear pattern by 16 hours APF (D,H,L). (A-D) Amos expression is detected throughout this time, but the expression is ectodermal from 0-4 hours APF, and then it co-labels with some of the precursors between 4-16 hours APF. Amos continues to be expressed in some cells at 16 hours APF, and these cells seem to be a mixture of ectodermal cells and precursors.

can be observed in large numbers of sensilla on the maturing pupal antenna. From their morphology, it is clear that the GFP-expressing subset are the sensilla trichodea and basiconica (Fig. 6G). This confirms that early SOPs form sensilla coeloconica whereas late SOPs produce sensilla trichodea and basiconica. Interestingly, these GFP-expressing sensilla differentiate late, because there is no overlap with the 22C10 marker until late in development (Fig. 6C,D). Thus, the timing of neuronal differentiation reflects the timing of SOP birth. These findings correlate with the differing effects of proneural genes on fasciculation as described above: the first-born *ato*-dependent cells organise the nerves, and the later *amos*-dependent ORNs follow passively.

Loss of olfactory precursors in *amos* mutants

Loss of SOPs is one of the defining characteristics of proneural gene mutations. We examined SOP formation in *amos* mutants relative to wild type by examining Sens expression. As expected from the expression analysis described above, the first two waves of SOP formation show little discernible difference in pattern between *amos* mutant antennal discs and wild-type discs (Fig. 5G,H). This is consistent with these early SOPs expressing and requiring *ato*, and they indeed express *ato* in a pattern indistinguishable from wild type (Fig. 5K). This shows that *ato* expression does not depend on *amos* function. After this, the later arising SOPs do not appear to form between rows of *ato*-dependent SOPs, corresponding to those cells shown to express *amos* (compare Fig. 5K with Fig. 4D). A few precursors do not express *ato*, and these may represent precursors of the ectopic bristles (Fig. 5K). This is supported by an analysis of Cut expression, which is the key molecular switch that must be activated to allow SOPs to take a bristle fate (Bodmer et al., 1987; Blochlinger et al., 1991), and whose expression correlates with bristle SOPs (Blochlinger et al., 1990). In the wild-type antenna, Cut is not expressed during olfactory SOP formation (Fig. 5M), although later it is expressed in differentiating cells of all olfactory sensilla (Fig. 6I and data not shown). This expression normally appears after 16 hours and does not overlap with Amos. In *amos* mutant antennae, expression begins earlier than normal in a subset of SOPs that appear to correspond to the ones identified above (Fig. 5N).

By 16 hours APF, there is a large loss of Sens staining in *amos* mutants (Fig. 5I,L). The remaining cells tend to be in clusters as would be expected for the early *ato*-dependent sensilla, but otherwise the identity of these cells cannot be determined. Detection of Amos protein in $amos^{1}$ mutant antennal discs also shows that although the Amos domains are still present, the deeper Amos/Sens-expressing nuclei are absent (Fig. 5J). Thus, at least a large number of *amos*-associated SOPs are not formed in the *amos* mutant.

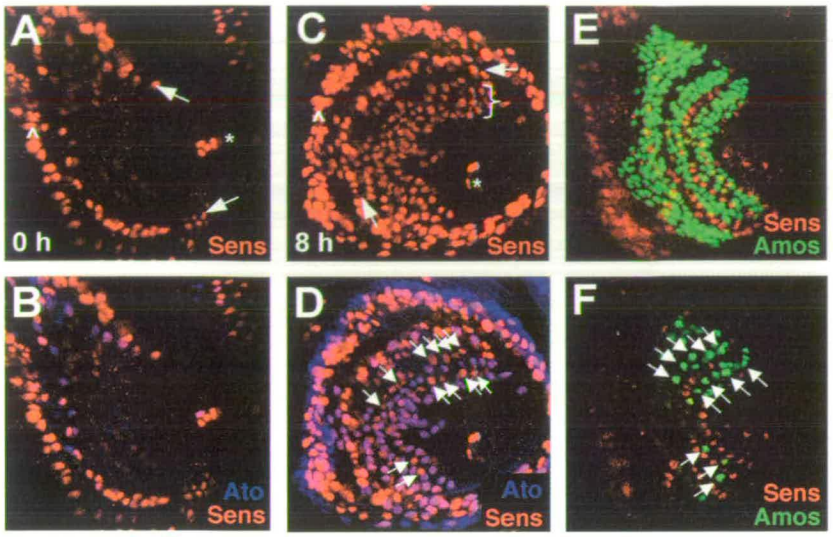


Fig. 4. Olfactory precursors and *amos/ato* expression. (A,B) *ato* is expressed in the wave 1 precursors. The arrows mark the ends of this semicircular line of olfactory precursors. *ato*-dependent arista precursors (*) and chordotonal precursors (caret) are also marked. (C,D) *ato* is expressed in the wave 2 olfactory precursors (bracket). It is not expressed in the third wave, the first cells of which can be seen between the Ato rows (arrows in D). (E,F) *amos* is expressed in a complementary way to *ato* in the third wave of precursors. (E) Confocal projection of a stack of images showing Amos detection in a similar disc to D. (F) Deep confocal section from E showing nuclei of *amos*-expressing precursors (arrows) underlying the main *amos* proneural expression domain. These correspond to the non-Ato-expressing precursors (arrows in D). (G) Schematic summary of *amos*- and *ato*-dependent precursor pattern at ~8 hours APF.

Expression of *amos* during sensillum development

The processes and lineages by which olfactory SOPs lead to the differentiated cells of the olfactory sensillum are not entirely known. The limited information available comes from analysis of the early wave of SOPs, which we have established represent the *ato*-dependent sensilla. After an SOP is selected there appears in its place a cluster of 2-3 cells expressing the A101 enhancer trap [the pre-sensillum cluster (PSC)]; this is apparently caused not by division of the SOP but perhaps by recruitment by the SOP (Ray and Rodrigues, 1995; Reddy et al., 1997), although the evidence for this is indirect. These PSC cells then divide to form the cells of the sensillum, including the outer support cells (hair and socket cells), inner support cells (sheath cells) and 1-4 neurons. For the early subset of SOPs, formation of the PSC occurs at a time in which *amos* is still expressed in the epithelial domains, and so *amos* could influence the development of these cells. Using A101 as a marker of the PSC cells, we determined that *amos* is not expressed in recognisable PSCs at 8 or 16 hours APF (Fig. 6E). Moreover, there is also no apparent co-labelling of Amos and Pros [a marker of one of the PSC cells (Sen et al., 2003)] (Fig. 6F). This suggests either that early PSC cells do not derive from *amos*-expressing cells or that *amos* is switched off rapidly when cells join a PSC.

The situation appears different for the cells derived from *amos*-dependent SOPs. Surprisingly at 24 hours and beyond, the *amos* enhancer drives GFP expression in most or all cells of the differentiating sensilla basiconica and trichodea (Fig. 6G): including most or all of the neurons (recognised by Elav expression; Fig. 6H); the sheath cell (recognised by Pros expression; Fig. 6H); and the outer support cells (recognised by the higher expression of Cut; Fig. 6I). This suggests that the late PSC cells do derive from *amos*-expressing cells and that activation of an enhancer within the 3.6 kb regulatory fragment (possibly separate from the SOP enhancer) is part of their specification process, although *amos* expression itself may not be long lived in these cells.

amos represses *scute* function

amos mutant antennae have Cut-expressing SOPs, but, although *cut* expression decides SOP subtype fate, it does not specify ectodermal cells as SOPs de novo. To investigate the involvement of other proneural genes, we first determined whether the bristles depended on *ato*, as it is expressed in close proximity to the emerging bristle SOPs. Clones of $amos^1$ mutant tissue were induced in ato^1 mutant antennae. In such clones, all olfactory sensilla were absent, as expected, but ectopic bristles were still formed (Fig. 1F). Therefore the bristles do not depend on *ato* function.

Cut expression normally follows from *ac/sc* proneural function, and so the ectopic bristle SOPs might depend on these proneural genes. Indeed, mutation of *ac* and *sc* greatly reduces the number of ectopic bristles in $amos^1$ flies ($In(1)sc^{10\text{-}1}/Y$; $amos^1/Df(2L)M36F\text{-}S6$ flies) (Fig. 1G and Table 2). By contrast, mutation of the non-proneural *ASC* gene *asense* (*ase*) had no effect alone (Table 2). This suggests that in the absence of *amos*, *ac/sc* function, to a large extent, causes the formation of bristle SOPs.

To determine how *amos* might normally repress bristle formation, we examined the pattern of *sc* mRNA in the pupal antenna. Significantly, a weak stripe of *sc* expression was observed in the wild-type antenna. (Fig. 7A). This stripe coincides with *amos* expression, and consists of ectodermal cells and SOPs (Fig. 7B,C). In the *amos* mutant antenna, *sc* mRNA expression was stronger and more clearly correlated with SOPs (Fig. 7C). This suggests that *sc* is expressed in olfactory regions of the wild-type antenna but that its function is repressed by the presence of *amos*. We therefore investigated *sc* functional activity in the antenna by analysing the expression of specific *sc* target genes as indicators of Sc protein function. Firstly, we examined Ac protein, whose expression is ordinarily activated by Sc function as a result of cross regulation (Gomez-Skarmeta et al., 1995). Ac protein is present in some SOPs in *amos* mutant antennae, but is not present in wild-type antennae (Fig. 7E,F). A similar result was observed for *sc-SOP-GFP*, which is a reporter gene construct that is directly activated by *sc* upon SOP formation (L. Powell and A.P.J., unpublished) (Culí and Modolell, 1998). This reporter showed GFP expression in some SOPs in *amos* mutant antennae but not in

wild-type antennae (data not shown). Finally, we examined *sc-E1-GFP*, a reporter gene construct comprising GFP driven solely by a *sc*-selective DNA binding site (L. Powell and A.P.J., unpublished) (Culí and Modolell, 1998). This reporter is invariably activated in all cells containing active Sc protein (including PNCs and SOPs) (L. Powell and A.P.J., unpublished). As with the other target genes, this reporter was only expressed in *amos* mutant antennae (Fig. 7G,H). Thus, we conclude that *sc* mRNA is expressed in the wild-type pupal antenna, and *amos* normally must repress either the translation of this RNA or the function of the Sc protein produced. This conclusion is supported by misexpression experiments. When *amos* is misexpressed in *sc* PNCs of the wing imaginal disc (*109-68Gal4*/UAS-*amos*) there is a dramatic reduction in bristle formation (Fig. 8A,B), even though endogenous *sc* RNA levels are unaffected (data not shown).

The transcription factor encoded by *lozenge* (*lz*) plays a number of roles in olfactory sensillum development, including activating *amos* expression (Goulding et al., 2000). Mutants therefore show a loss of many *amos*-dependent sensilla. Interestingly flies mutant for both *lz* and *amos* (lz^{34}; $amos^{1}/Df(2L)M36F\text{-}S6$) have third antennal segments that bear only sensilla coeloconica, and so the ectopic bristles of *amos* mutants are dependent on *lz* function (Table 2, Fig. 1H). Correlating with this, the expression of *sc* mRNA in the third antennal segment was much reduced in a *lz* mutant compared with wild type (Fig. 7D). Thus, *lz* appears at least partly responsible for the expression of *sc* in the antenna.

Discussion

We show definitively that *amos* is the proneural gene for the precursors of two classes of olfactory sensilla. These precursors are absent in *amos* mutants, resulting in highly defective antennae lacking all sensilla basiconica and trichodea. Unusually, this is not the only phenotype of *amos* mutants. Unique among *Drosophila* proneural genes, mutation of *amos* results in the appearance of new sense organs: mechanosensory bristles are now formed on the third antennal segment. We provide evidence that *amos* must normally repress *sc*-promoted bristle specification in addition to promoting olfactory neurogenesis. Significantly, inhibitory interactions between bHLH genes have recently been reported during mouse neurogenesis, where discrete domains of bHLH transcription factor expression are set up partly by mutual cross-inhibition combined with autoregulation (Gowan et

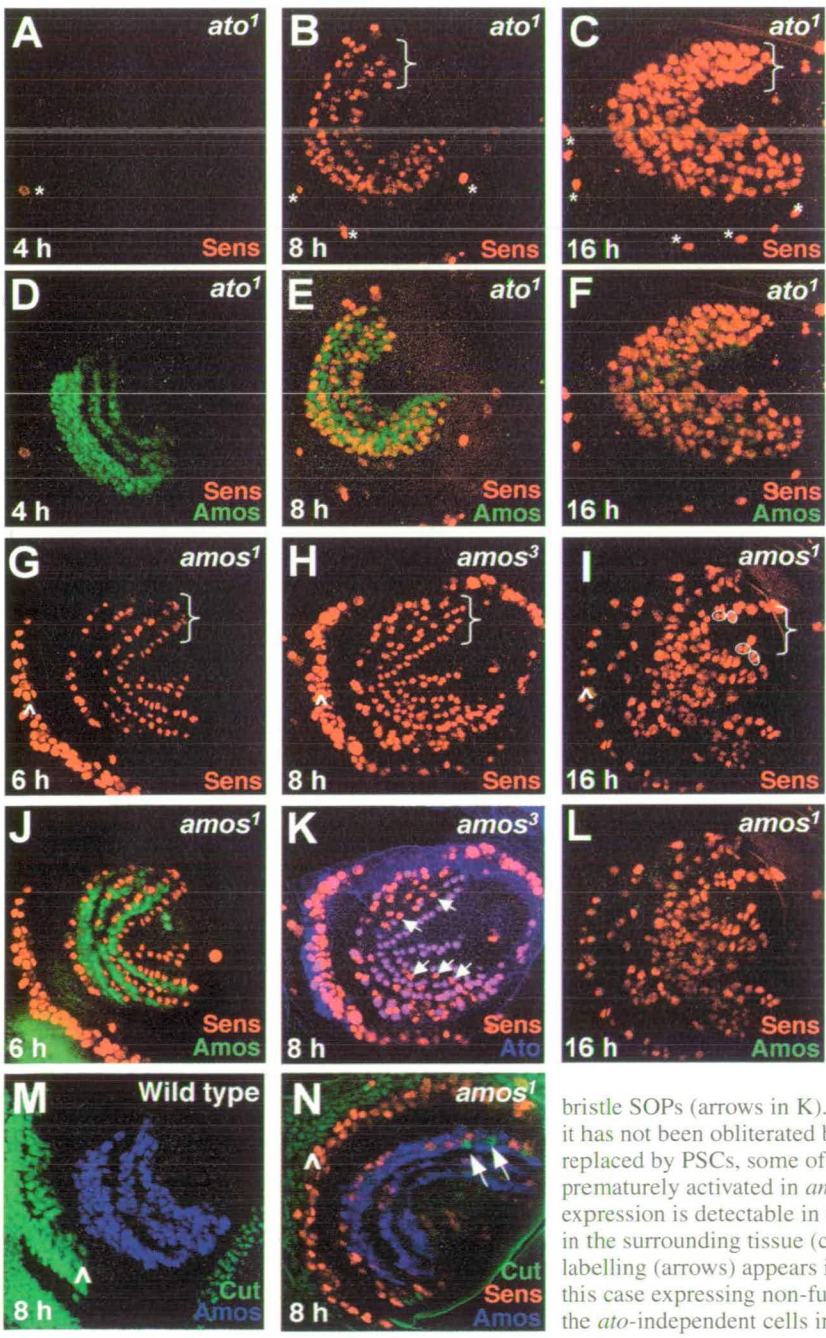


Fig. 5. Olfactory precursors in *amos* and *ato* mutants. These discs should be compared with the corresponding wild-type discs in Figs 3 and 4. (A,D) The early precursors are specifically lost in *ato* mutants. The remaining olfactory precursors correspond to the third wave (B,C) and align very closely with the *amos* expression domains (E,F). In the second segment, the chordotonal precursors are also missing and only a few bristle precursors remain (*, A-C). (G-L) The late precursors are specifically lost in *amos* mutants. (G,J) Early precursor pattern resembles wild type, with mutant $Amos^{1}$ protein detectable between the rows of precursors (brackets). Caret in G indicates chordotonal precursors. (H,K) At 8 hours APF, the pattern remains unchanged as the third wave SOPs are not formed (c.f. Fig. 4C). These early precursors mostly express Ato, although a number of non-Ato expressing SOPs appear between the early rows, which could correspond to the bristle SOPs (arrows in K). (I,L) The early pattern is still apparent at 16 hours APF as it has not been obliterated by the third wave of SOPs (the early SOPs have now been replaced by PSCs, some of which are ringed). (M,N) Cut expression appears prematurely activated in *amos* mutants. (M) Wild type at 8 hours APF. No Cut expression is detectable in the third segment SOPs; however, Cut stains very strongly in the surrounding tissue (caret). (N) $amos^{1}$ mutant at 8 hours APF. Some Cut labelling (arrows) appears in SOPs derived from the Amos-expressing domains (in this case expressing non-functional $Amos^{1}$ protein). These cells seem to correspond to the *ato*-independent cells in K.

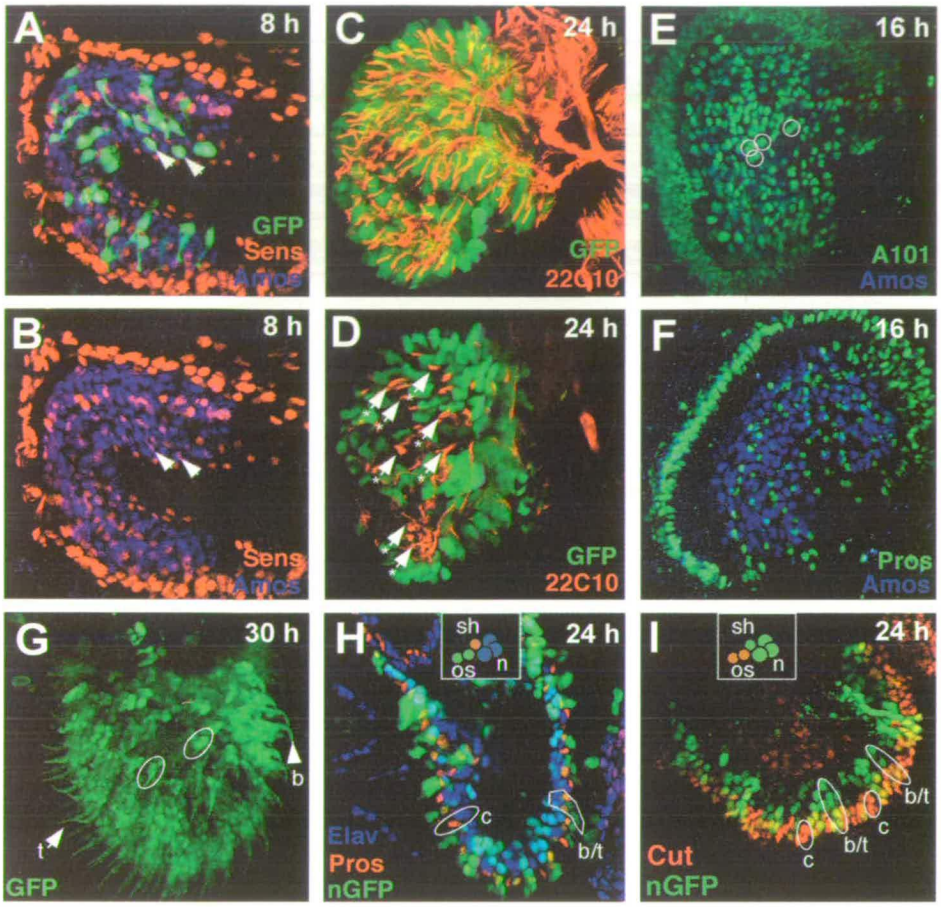


Fig. 6. Fate of *amos*-expressing cells during olfactory development. (A-D,G) Activity of an *amos* SOP enhancer driving GFP expression. (A,B) At 8 hours APF, GFP can be detected in the third wave of olfactory precursors, some co-labelled SOPs are indicated by arrows (co-labelled with Sens and Amos in the separation in B). (C,D) The *amos-GFP* expressing cells contribute to late differentiating sensilla, as shown by lack of co-labelling with a neuronal marker (22C10) at 24 hours APF. C is a projection of many sections whereas D is a confocal section with some of the differentiating neurons marked by asterisks, these do not express GFP. (E,F) Later expression of *amos* does not correspond to PSC cells. (E) At 16 hours APF, Amos expression is fading, but there is no overlap with A101 β-galactosidase expression in the PSCs, some of which are ringed. (F) There is no overlap of Amos expression with that of Pros, a marker of one of the PSCs. (G) *amos-GFP* construct at 30 hours APF: a large number of sensilla retain GFP. Protein appears to be in sensillar groups (as indicated by rings), and includes the outer support cells, so that sensilla trichodea (t) and basiconica (b) can clearly be discerned. (H,I) Analysis of *amos-GFP* in confocal sections of antennae at 24 hour APF relative to the component cells of the sensilla (see insets). n, neuron; sh, sheath; os, outer support cells. *amos-GFP* labels rows of cells corresponding to each sensillum basiconicum or trichodeum (some are ringed), whereas presumptive coeloconica (c) do not express GFP. (H) GFP is expressed in neurons (marked by Elav) and sheath cells (marked by Pros). (I) GFP is expressed in outer support cells (marked by stronger expression of Cut).

al., 2001; Nieto et al., 2001). As with *amos*, cross-inhibition occurs between members of different bHLH families: *Mash1* (ASC homologue), *Math1* (*ato* homologue), and *neurogenin1* (*tap* homologue).

How proneural genes determine neuronal subtype

On misexpression evidence, we have argued that neuronal subtype specification involves repression of bristle fate by *ato* during chordotonal SOP formation (Jarman and Ahmed, 1998) and by *amos* during olfactory precursor formation (Goulding et al., 2000). In this light, the ectopic bristles in *amos* mutants are of significant interest. They represent the first loss-of-function evidence that an *ato*-type proneural gene suppresses bristle fate during the normal course of its function. However, how this relates to *amos* function is complex. In misexpression experiments, bristle suppression by *amos* is most strongly observed using a PNC- and SOP-specific Gal4 driver line (Goulding et al., 2000) (this report). Yet paradoxically, misexpression of *amos* more generally in the ectoderm, but only weakly in SOPs, yields dramatically different results: in such cases *amos* produces ectopic bristles very efficiently (Huang et al., 2000; Lai, 2003; Villa Cuesta et al., 2003). This bristle formation does not require the function of endogenous *ac/sc* genes (Lai, 2003), but probably reflects the intrinsic SOP-specifying function of *amos* in situations that are not conducive to its subtype-specifying (and bristle suppressing) function. It appears therefore that bristle suppression particularly requires *amos* expression in SOPs.

What does *amos* repress in the antenna? It appears that *sc* is expressed within the wild-type *amos* expression domain during olfactory SOP formation. Clearly *amos* must prevent the function of *sc*, as *sc* expression in ectoderm usually results in bristle specification. It may be significant that some of the *sc* RNA is in olfactory SOPs in the wild-type antenna, suggesting that the SOP may be a major location of repression by *amos*, as indicated by misexpression experiments. However, some bristle formation is maintained in *ac/sc; amos* mutants. This may be due to redundancy with other genes in the ASC: certainly wild-type bristle formation outside the antenna is not completely abolished in the absence of *ac/sc* (A.P.J., unpublished). An alternative possibility is that some bristle SOPs result from other proneural-like activity in the antenna. Direct proneural activity of *lz* is a possibility, although misexpression of *lz* elsewhere in the fly (using a *hs-lz* construct) is not sufficient to promote bristle formation (P.I.z.L., unpublished).

The *amos*^{2} hypomorph appears to represent a different situation. In such flies, a number of *amos*-dependent SOPs appear to have mixed olfactory/bristle fate. This suggests that on occasions the mutant Amos^{2} protein is able to specify SOPs, but is less able to impose its subtype function (and so this, to some extent, resembles more the outcome of some misexpression experiments). *amos*^{2} may therefore be a useful

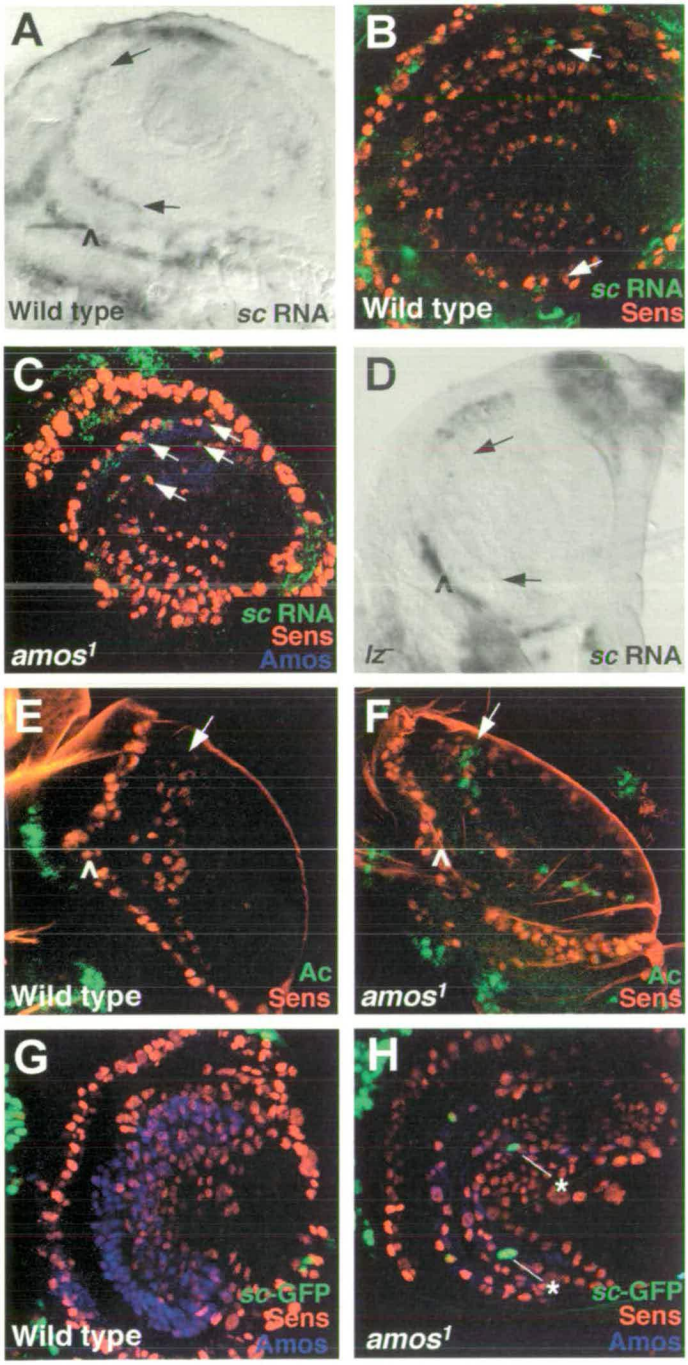


Fig. 7. Expression of *sc* and *sc* target genes in the antenna. (A-D) *sc* mRNA detected by in situ hybridisation. (A) Wild type, with *sc* expressed not only in the second antennal segment (caret) but also in the third segment (arrows). (B) Wild type, with sc RNA detected by immunofluorescence (green). (C) *amos*^{1} mutant. *sc* mRNA is increased and is present in SOPs (arrows). (D) The second segment *sc* expression is reduced in *lz*^{34} mutants. (E,F) Ac expression is present in some SOPs in *amos* mutants. (E) Wild type at 8 hours APF, showing very little Ac expression in the third segment (first precursor wave marked by arrow) (some is visible in the second segment; caret). (F) *amos*^{3} mutant at 8 hours APF, showing some Ac expression in second segment (arrow). (G,H) GFP expression from *sc-E1-GFP* reporter transgene. (G) Wild type, showing no expression in third segment. (H) *amos*^{3} mutant showing expression (*).

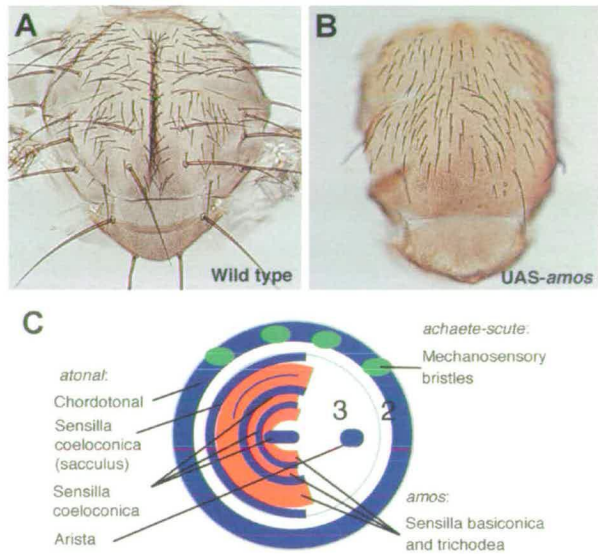


Fig. 8. *amos* misexpression represses bristle formation. (A) Wild type dorsal thorax. (B) Dorsal thorax from *109-68Gal4*/UAS-*amos* fly. *amos* misexpression driven in *sc* PNCs by this driver line results in loss of many bristles (mainly the large macrochaetae). (C) Summary of proneural functions in antenna. Diversity of sense organs laid down by function of three proneural gene systems. Blue, *atonal*; red, *amos*; green, *achaete/scute*.

tool for exploring these two functions. For example, if subtype specification requires interaction of Amos with protein co-factors (Jarman and Ahmed, 1998; Brunet and Ghysen, 1999; Hassan and Bellen, 2000), then these interactions may be specifically impaired in the *amos*^{2} mutant.

Because the proneural proteins are normally transcriptional activators, it is unlikely that Amos/Ato proteins directly inhibit gene expression during bristle suppression (Jarman and Ahmed, 1998). The presence of *sc* RNA in *amos*-expressing cells in the wild-type antenna is consistent with this. The involvement of protein interactions is to be suspected. An interesting parallel is found in vertebrates, where neurogenin1 promotes neurogenesis and inhibits astrocyte differentiation (Nieto et al., 2001). The glial inhibitory effect could be separated from the neurogenesis promoting effect: whereas neurogenesis promotion depends on DNA binding and activation of downstream target genes, astrocyte differentiation was inhibited through a DNA-independent protein-protein interaction with CBP/p300 (Sun et al., 2001; Vetter, 2001). In the case of *amos*, an interesting possibility is that inhibition of bristle formation may involve the sequestering of Sc protein by Amos protein. As discussed above, such a mechanism would have to be sensitive to the level or pattern of *amos*, as general misexpression does not mimic this activity.

Comparison of *amos* and *ato* as olfactory proneural genes

Apart from giving rise to separate classes of olfactory precursor, there are interesting differences in the way that *ato* and *amos* are deployed in the antenna. We characterised three waves of olfactory precursor formation (Fig. 4G). The first and second waves are well defined, giving rise to well-patterned

sensilla coeloconica of the sacculus and the antennal surface, respectively. These precursors express and require *ato*. The third wave of precursors is much more extensive and has little obvious pattern; it gives rise to the much more numerous sensilla basiconica and trichodea. This wave expresses and requires *amos*. For the early waves, *ato* is expressed according to the established paradigm: it is expressed in small PNCs, each cluster giving rise to an individual precursor (Gupta and Rodrigues, 1997). The pattern of the PNCs is very precise and prefigures the characteristic pattern of precursors. *amos* expression is dramatically different. It is expressed in large ectodermal domains for an extended period of time. Densely packed precursors arise from this domain continuously without affecting the domain expression. This shows that singling out does not necessarily require shut down of proneural expression, and therefore has implications for how singling out occurs. In current models, it is assumed that PNC expression must be shut down to allow an SOP to assume its fate. The *amos* pattern better supports the idea that a mechanism of escaping from or becoming immune to lateral inhibition is more likely to be important generally. One prediction would be that *amos* and *ato* (and *ac/sc*) differ in their sensitivities to Notch-mediated lateral inhibition, a situation that has been noted for mammalian homologues (Lo et al., 2002).

Why are the proneural genes deployed so differently? One possibility is simply that there are very many more sensilla basiconica and trichodea than coeloconica. All the coeloconica precursors can be formed by *ato* action in a precise pattern in two defined waves. This would not be possible for the large number of basiconica and trichodea precursors, and so precursor selection has been modified for *amos*. Indeed, *amos* appears to be a particularly 'powerful' proneural gene when misexpressed (Lai, 2003; Villa Cuesta et al., 2003). This may make *amos* a useful model of other neural systems in which large numbers of precursors must also be selected.

For most insects, the antenna is the major organ of sensory input. It is not only the site of olfaction, but also of thermoreception, hygroreception, vibration detection and proprioception, as well as of touch. Patterning the sensilla is therefore complex and three types of proneural gene are heavily involved to give different SOPs (Fig. 7G). It is clear that the study of antennal sensilla will provide a useful model for exploring the fate determining contribution of intrinsic bHLH protein specificity and extrinsic competence factors.

We wish to thank Anindya Sen for help and discussions. We thank Hugo Bellen for the anti-Sens antibodies, and the Developmental Biology Hybridoma Bank, Iowa for other antibodies. This work was supported by a Wellcome Trust project grant (055851) and Senior Research Fellowship to A.P.J. (042182).

References

Bertrand, N., Castro, D. S. and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517-530.

Blochlinger, K., Bodmer, R., Jan, L. Y. and Jan, Y. N. (1990). Patterns of expression of Cut, a protein required for external sensory organ development in wild-type and *cut* mutant *Drosophila* embryos. *Genes Dev.* **4**, 1322-1331.

Blochlinger, K., Jan, L. Y. and Jan, Y. N. (1991). Transformation of sensory organ identity by ectopic expression of Cut in *Drosophila*. *Genes Dev.* **5**, 1124-1135.

Bodmer, R., Barbel, S., Shepherd, S., Jack, J. W., Jan, L. Y. and Jan, Y. N. (1987). Transformation of sensory organs by mutations of the *cut* locus of *D. melanogaster*. *Cell* **51**, 293-307.

Brunet, J.-F. and Ghysen, A. (1999). Deconstructing cell determination: proneural genes and neuronal identity. *BioEssays* **21**, 313-318.

Carlson, J. R. (1996). Olfaction in *Drosophila*: from odor to behavior. *Trends Genet.* **12**, 175-180.

Culí, J. and Modolell, J. (1998). Proneural gene self-stimulation in neural precursors: an essential mechanism for sense organ development that is regulated by *Notch* signalling. *Genes Dev.* **12**, 2036-2047.

Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., FerresMarco, D., Beamonte, D. and Modolell, J. (1995). Cis-regulation of *achaete* and *scute*: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev.* **9**, 1869-1882.

Goulding, S. E., zur Lage, P. and Jarman, A. P. (2000). *amos*, a proneural gene for *Drosophila* olfactory sense organs that is regulated by *lozenge*. *Neuron* **25**, 69-78.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R. and Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron* **31**, 219-232.

Gupta, B. P. and Rodrigues, V. (1997). *atonal* is a proneural gene for a subset of olfactory sense organs in *Drosophila*. *Genes Cells* **2**, 225-233.

Hassan, B. A. and Bellen, H. J. (2000). Doing the MATH: is the mouse a good model for fly development? *Genes Dev.* **14**, 1852-1865.

Hassan, B. A., Bermingham, N. A., He, Y., Sun, Y., Jan, Y. N., Zoghbi, H. Y. and Bellen, H. J. (2000). *atonal* regulates neurite arborization but does not act as a proneural gene in the *Drosophila* brain. *Neuron* **25**, 549-561.

Huang, M. L., Hsu, C. H. and Chien, C. T. (2000). The proneural gene *amos* promotes multiple dendritic neuron formation in the *Drosophila* peripheral nervous system. *Neuron* **25**, 57-67.

Jarman, A. P. and Ahmed, I. (1998). The specificity of proneural genes in determining *Drosophila* sense organ identity. *Mech. Dev.* **76**, 117-125.

Jarman, A. P., Grau, Y., Jan, L. Y. and Jan, Y. N. (1993). *atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* **73**, 1307-1321.

Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y. and Jan, Y. N. (1994). *atonal* is the proneural gene for *Drosophila* photoreceptors. *Nature* **369**, 398-400.

Jhaveri, D., Sen, A., Reddy, G. V. and Rodrigues, V. (2000a). Sense organ identity in the *Drosophila* antenna is specified by the expression of the proneural gene *atonal*. *Mech. Dev.* **99**, 101-111.

Jhaveri, D., Sen, A. and Rodrigues, V. (2000b). Mechanisms underlying olfactory neuronal connectivity in *Drosophila* – the *atonal* lineage organizes the periphery while sensory neurons and glia pattern the olfactory lobe. *Dev. Biol.* **226**, 73-87.

Johnston, L. A., Ostrow. B. D., Jasoni, C. and Blochlinger, K. (1998). The homeobox gene *cut* interacts genetically with the homeotic genes *proboscipedia* and *Antennapedia*. *Genetics* **149**, 131-142.

Lai, E. C. (2003). *Drosophila Tufted* is a gain-of-function allele of the proneural gene *amos*. *Genetics* **163**, 1413-1425.

Lindsley, D. L. and Zimm, G. G. (1992). *The Genome of* Drosophila melanogaster. San Diego: Academic Press.

Lo, L., Dormand, E., Greenwood, A. and Anderson, D. J. (2002). Comparison of generic neuronal differentiation and neuron subtype specification functions of mammalian *achaete-scute* and *atonal* homologues in cultured neural progenitor cells. *Development* **129**, 1553-1567.

Newsome, T. P., Asling, B. and Dickson, B. J. (2000). Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851-860.

Nieto, M., Schuurmans, C., Britz, O. and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. *Neuron* **29**, 401-413.

Nolo, R., Abbott, L. A. and Bellen, H. J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102**, 349-362.

Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J. and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. *Genes Dev.* **16**, 324-338.

Ray, K. and Rodrigues, V. (1995). Cellular events during development of the olfactory sense organs in *Drosophila melanogaster*. *Dev. Biol.* **167**, 426-438.

Reddy, G. V., Gupta, B., Ray, K. and Rodrigues, V. (1997). Development of the *Drosophila* olfactory sense organs utilizes cell-cell interactions as well as lineage. *Development* **124**, 703-712.

Sen, A., Reddy, G. V. and Rodrigues, V. (2003). Combinatorial expression of Prospero, Seven-up and Elav identifies progenitor cell types during sense-organ differentiation in the *Drosophila* antenna. *Dev. Biol.* **254**, 79-92.

Shanbhag, S. R., Müller, B. and Steinbrecht, R. A. (1999). Atlas of olfactory organs of *Drosophila melanogaster*. 1. Types, external organization, innervation and distribution of olfactory sensilla. *Int. J. Insect Morph. Embryol.* **28**, 377-397.

Sun, Y., Jan, L. Y. and Jan, Y. N. (1998). Transcriptional regulation of *atonal* during development of the *Drosophila* peripheral nervous system. *Development* **125**, 3731-3740.

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G. and Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365-376.

Vetter, M. (2001). A turn of the helix: preventing the glial fate. *Neuron* **29**, 559-562.

Villa Cuesta, E., de Navascues, J., Ruiz Gomez, M., del Corral, R. D., Dominguez, M., de Celis, J. F. and Modolell, J. (2003). *Tufted* is a gain-of-function allele that promotes ectopic expression of the proneural gene *amos* in *Drosophila*. *Genetics* **163**, 1403-1412.

The *Drosophila* proneural gene *amos* promotes olfactory sensillum formation and suppresses bristle formation

Petra I. zur Lage, David R. A. Prentice, Eimear E. Holohan and Andrew P. Jarman*

The Wellcome Trust Centre for Cell Biology, Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

*Author for correspondence (e-mail: andrew.jarman@ed.ac.uk)

Accepted 20 June 2003

Development 130, 4683-4693

doi:10.1242/dev.00680

Summary

Proneural genes encode basic-helix-loop-helix (bHLH) transcription factors required for neural precursor specification. Recently *amos* was identified as a new candidate *Drosophila* proneural gene related to *atonal*. Having isolated the first specific *amos* loss-of-function mutations, we show definitively that *amos* is required to specify the precursors of two classes of olfactory sensilla. Unlike other known proneural mutations, a novel characteristic of *amos* loss of function is the appearance of ectopic sensory bristles in addition to loss of olfactory sensilla, owing to the inappropriate function of *scute*. This supports a model of inhibitory interactions between proneural genes, whereby *ato*-like genes (*amos* and *ato*) must suppress sensory bristle fate as well as promote alternative sense organ subtypes.

Key words: Proneural, bHLH, *Drosophila*, *amos*, Neurogenesis, Gene regulation

Introduction

The sequence and structure of the bHLH domain is highly conserved, and yet transcription factors of this family play a variety of roles in neurogenesis in a range of organisms (Bertrand et al., 2002). These roles include conferring neuronal competence, directing neural precursor specification, directing neuronal subtype specification and triggering neuronal differentiation. Dissecting bHLH gene functions and interactions is an important and challenging task, and the *Drosophila* PNS provides a good model in which to do this. Here, proneural bHLH genes are required for sense organ precursor (SOP) specification (Hassan and Bellen, 2000). These genes include *achaete* (*ac*) and *scute* (*sc*), from the *Achaete-scute Complex* (*ASC*), and *atonal* (*ato*), as well as the candidate proneural gene *amos* (Hassan and Bellen, 2000). These proneural proteins seem to combine two functions: promoting SOP specification, and providing these SOPs with information concerning neuronal subtype (Jarman and Ahmed, 1998). It is thought that vertebrate proneural gene homologues also have functions in neural progenitor specification and neural subtype identity (Hassan and Bellen, 2000; Bertrand et al., 2002).

bHLH functions depend on both intrinsic protein properties and extrinsic factors (Bertrand et al., 2002). Comparisons of protein capabilities, particularly by assaying the effect of misexpression on neural development, have shown evidence for intrinsic differences between closely related bHLH proteins, suggesting that they regulate distinct target genes (Jarman and Ahmed, 1998). However, bHLH protein specificity is also very dependent on extrinsic modifying factors. Ato has been well characterised and illustrates well the complexity of defining the intrinsic specificity of proneural proteins. In most of the developing ectoderm, Ato is required for chordotonal (stretch receptor) SOP specification. Ectopic expression of *ato* leads to ectopic chordotonal SOP formation (Jarman et al., 1993). In this property, it differs from Ac and Sc, which are necessary and sufficient for external sense organ (bristle) SOPs. This points to intrinsic differences in protein properties. However, the function of Ato is clearly also very context dependent. In addition to specifying chordotonal organs, Ato is also required for R8 photoreceptors in the eye (Jarman et al., 1994), and for one subset of olfactory sensilla (sensilla coeloconica) in the antenna (Gupta and Rodrigues, 1997). Moreover, in a group of CNS neurons, Ato regulates neurite arborization (Hassan et al., 2000). It is not known how the response to Ato is modified in these different regions.

We have argued for a specific mechanism by which proneural proteins specify neural subtype: SOPs may be biased to become external sense organs and, consequently, Ac/Sc promotes a default neural fate, whereas Ato must actively impose alternative neural fates (Jarman and Ahmed, 1998). This idea is based on two apparently paradoxical outcomes of misexpression experiments. Under certain very defined conditions, *ato* misexpression can transform *existing* bristle SOPs to chordotonal organs, thereby revealing an intrinsic ability of Ato (Jarman and Ahmed, 1998). However, in most contexts, *ato* misexpression induces a mixture of ectopic chordotonal and bristle SOPs (Jarman et al., 1993), suggesting that in many circumstances Ato can specify SOPs but may often fail to provide subtype information. This suggests that the two proneural roles are separable in misexpression studies, and it also gives the appearance that Ato function is more sensitive to cell context than is Ac/Sc function. Similar controlled misexpression data for vertebrate bHLH genes have recently been reported, which support an entirely analogous situation in which *neurogenin* (*ato* homologue) is more context

sensitive than *Mash1* (*ac/sc* homologue) (Lo et al., 2002) (see also Parras et al., 2002). But teasing out these functions is complicated and misexpression data could be misleading. There is no corroborative evidence from loss-of-function mutations in *Drosophila* as known proneural mutations always cause loss of SOP subsets, and so questions concerning the neural identity of SOPs are hard to approach through loss-of-function studies.

Recently, we and others described a new candidate proneural gene, *amos* (Goulding et al., 2000; Huang et al., 2000). Amos protein possesses a very similar bHLH domain to that of Ato, suggesting there may be functional similarities with Ato that set this gene pair apart from *ac/sc*. We provided strong but indirect evidence that *amos* is the proneural gene for the *ato*-independent classes of olfactory sensillum (sensilla basiconica and trichodea) (Goulding et al., 2000). Here, we report a detailed analysis of *amos* expression and function, including the first isolation and characterisation of specific *amos* mutations. We find that Amos protein is expressed in, and is required for, a late wave of olfactory SOPs in the antenna. These are the precursors for sensilla basiconica and trichodea, proving that *amos* is the proneural gene for these subtypes. However, an unexpected aspect of the mutant phenotype was the appearance of ectopic sensory bristles in place of the olfactory sensilla on the antenna. This replacement of sense organs rather than complete absence is unprecedented for a *Drosophila* proneural gene mutation. Our analysis suggests that loss of *amos* results in loss of olfactory sensilla and concomitant derepression of *ac/sc* leading to formation of external sense organ SOPs. This phenotype supports the argument that the *ato*-like proneural genes (*amos* and *ato*) suppress external sense organ fate as well as promote alternative neural fates.

Materials and methods

Fly stocks

Wild-type flies are generally Oregon R and pr^1, as appropriate. UAS-*amos* is described by Goulding et al. (Goulding et al., 2000). sc^{10-1}, ase^1 and lz^{34} are described by Lindsley and Zimm (Lindsley and Zimm, 1992). Deficiencies and mutants were obtained from the Umea stock centre. For *ato/amos* double-mutant analysis, $amos^1$ clones were induced in an ato^1 background by the FLP/FRT method using *eyelessFLP* (Newsome et al., 2000). The flies had the following genotype: *y w eyFLP; amos*^{1} *pr*^{1} *FRT40A/2×nlsGFP FRT40A; ato*^{1}. Clones were recognised by their sensillum phenotype.

Mutagenesis

$amos^1$ was isolated in an F2 screen for mutations that failed to complement a deficiency of the *amos* region [*Df(2L)M36F-S6* (Goulding et al., 2000)]. pr^1 male flies were mutagenised with 25-30 mM EMS. Mutagenised lines were collected over a CyO balancer and individually tested for complementation with *Df(2L)M36F-S6*. 4500 mutagenised lines were screened. $amos^2$ and $amos^3$ were isolated in a subsequent F1 screen of 25,000 flies using $amos^1$. PCR isolation of the ORFs and sequencing were by standard techniques.

Amos enhancer construct

A 3.6 kb fragment upstream of the *amos* start site was amplified by PCR and cloned into the transformation vector pTLGal4 (a gift of B. Hassan). Transformant flies were made by microinjection into syncytial blastoderm embryos. These were crossed to UAS-*GFP* or UAS-*nlsGFP* lines for assessment of enhancer activity.

Immunohistochemistry

Antibody staining of pupal antennae was carried out as previously described (Goulding et al., 2000). Pupae were staged by collecting at the time of puparium formation and then ageing on moist filter paper at 25°C before dissection. Antibodies used were: Cut (1:100), Ac (1:50), 22C10 (1:200) and Elav (1:200) (all from the Developmental Biology Hybridoma Bank, Iowa); Sens (1:6250) (Nolo et al., 2000); and Pros (1:200). Anti-Amos antibodies were raised in rabbits, using full-length His_6-tagged Amos protein expressed in *E. coli*, and purified by adsorption to nickel-agarose under denaturing conditions. Anti-Amos antibodies were used at 1:1250 after pre-adsorption against wild-type embryos. RNA in situ hybridisation was done according to standard protocols using digoxigenin-labelled *sc* cDNA. RNA/protein double labellings were carried out by initially detecting RNA using anti-digoxigenin-POD and an Alexa Fluor 488 tyramide substrate (Molecular Probes), followed by antibody staining. Microscopy analysis was carried out using an Olympus AX70 or Leica LCS-SP system.

Results

amos mutations result in loss of olfactory sensilla and the appearance of mechanosensory bristles

We generated three mutant alleles of *amos* in an EMS screen (Table 1). $amos^1$ is predicted to result in a protein truncation that removes the second half of bHLH helix 2 and the C-terminal region thereafter. $amos^2$ is a missense mutation that changes a Ser to an Asn in helix 1 of the bHLH domai . This position is not part of the bHLH core consensus sequence and is not predicted to affect directly DNA binding or dimerisation. Moreover, Asn is found in this position in the *ato* bHLH domain, and so the effect of this mutation would be predicted to be mild. $amos^3$ contains a 230 bp deletion within the ORF, which also causes a frame-shift that brings a spurious downstream stop codon in frame. This allele gives a predicted peptide of 74 amino acids, of which only the first 30 are shared with *amos*. It therefore lacks the entire bHLH domain and is likely to be a null.

Consistent with previous RNAi experiments (Huang et al., 2000), $amos^1$ mutant embryos lack two dorsal sensory neurons per segment, including the dorsal bipolar dendritic neuron (D.R.A.P. and A.P.J., unpublished). Nevertheless, all *amos* alleles are adult viable as homozygotes and hemizygotes. The antennae of mutant adult flies were mounted and examined by light microscopy in order to quantify the number and type of olfactory sensilla. Compared with wild-type (Fig. 1A and Table 2) (Carlson, 1996), *amos* mutant antennae carried dramatically reduced numbers of sensilla and, as a consequence, the third segment is significantly smaller (Fig. 1C,D). In particular,

Table 1. Molecular basis of *amos* mutations

Allele	DNA	Protein	Predicted effect	Phenotype
$amos^1$	C_{550}>T_{550}	Q_{184}>Stop	Truncates the bHLH domain	Strong hypomorph/null
$amos^2$	G_{458}>A_{458}	S_{153}>N_{153}	Substitution in bHLH domain	Moderate hypomorph
$amos^3$	230 bp deletion+frameshift		Severely truncated protein	Null?

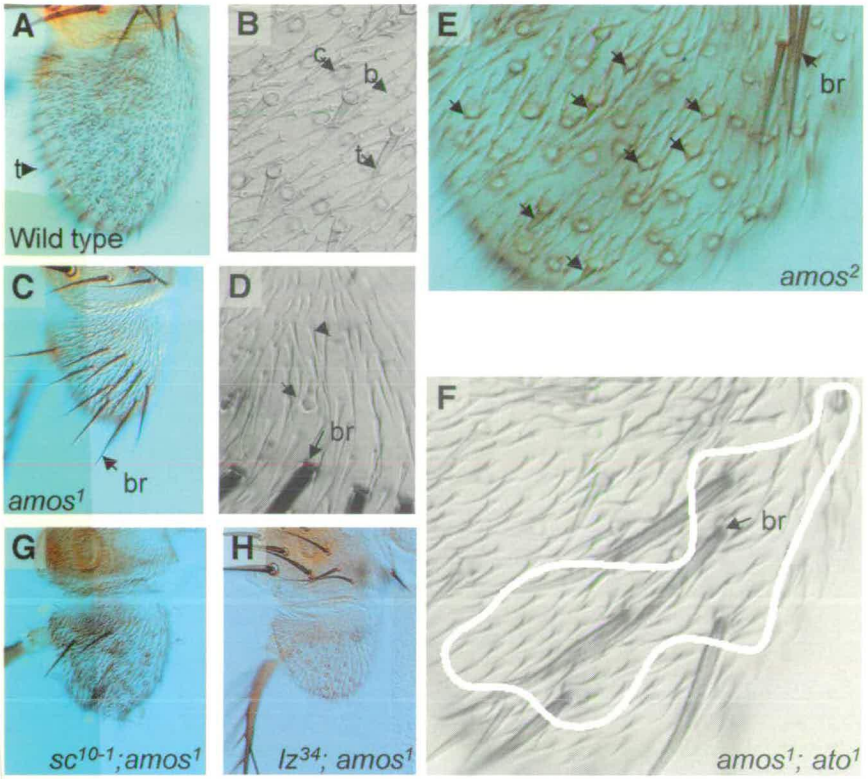


Fig. 1. Antennal defects in *amos* mutants. Third antennal segments are shown. (A) Wild-type antenna, with trichodea indicated (t). (B) Higher power view with examples marked of sensilla coeloconica (c), basiconica (b) and trichodea (t). (C) *amos*^{1}*/Df(2L)M36-S6*. Third segment is reduced because of missing basiconica and trichodea. Ectopic mechanosensory bristles are indicated (br). (D) Higher power view showing an abnormal domed sensillum (arrow) and a normal sensillum coeloconicum (arrowhead). Sensilla are very sparse. (E) *amos*^{2} with numerous abnormal sensilla (arrows). (F) Double mutant for *amos* and *ato* (clone of *amos*^{1} tissue in *ato*^{1} fly: *eyFLP; FRT-amos*^{1}*/FRT-nlsGFP; ato*^{1}). The clone patch contains no sense organs except bristles. (G) Double mutant for *amos* and *sc*$^{10\text{-}1}$ (which removes both *ac* and *sc* function). The extra bristles of *amos*^{1} are largely dependent on *ac/sc* function. (H) Double mutant for *amos* and *lz*. The extra bristles of *amos*^{1} are absent and therefore depend on *lz* function.

sensilla basiconica and trichodea were completely absent in the probable genetic nulls, *amos*^{1} and *amos*^{3}, whereas sensilla coeloconica appeared unaffected (Table 2). These phenotypes support the assertion that *amos* is the proneural gene for sensilla basiconica and trichodea, whereas *ato* is the proneural gene for sensilla coeloconica. However, mutant antennae exhibit a further unexpected phenotype. In wild type, the third segment bears only olfactory sensilla; in *amos* mutant flies, this segment bears a number of ectopic external sensory bristles and other abnormally structured sensilla (Fig. 1C,D and Table 2). These bristles do not have bracts (unlike bristles on the leg), and so this phenotype does not represent a transformation of antenna to leg (c.f. Johnston et al., 1998). These phenotypes are highly unusual as a characteristic of all other loss-of-function proneural gene mutations is that they cause the loss of sense organ subsets without the concomitant appearance of new or abnormal sensory structures. Therefore, the *amos* null phenotype is unique for a *Drosophila* proneural gene.

Given its subtle molecular basis, the putative hypomorph *amos*^{2} has a surprisingly strong phenotype: it has no sensilla trichodea, and sensilla basiconica are reduced very substantially (Table 2). There are also fewer ectopic bristles than in the null alleles, but there are many sensilla of unusual morphology. In the case of this allele, these seem to represent intermediates between sensilla basiconica/trichodea and external sense organs (Fig. 1E).

Late pupal antennae were stained with a sensory neuron marker, MAb22C10, to visualise olfactory receptor neurons (ORNs). Olfactory sensilla are innervated by multiple sensory neurons (Shanbhag et al., 1999), which can be seen as groups in the wild-type antenna (Fig. 2A). *amos* mutant antennae have many fewer neuronal groups, corresponding in number to the sensilla coeloconica and the bristles (Fig. 2B). There are instances of sensilla innervated by a single neuron, which appear to correspond to the ectopic bristles (Fig. 2C,D). In wild-type flies, ORN axons form three olfactory nerves leading to the antennal lobe of the brain (Jhaveri et al., 2000b) (Fig. 2E). In *amos* mutant antennae, all three antennal nerves are still

Table 2. Sensillum numbers on adult antennae

Genotype	Basiconica	Trichodea	Coeloconica	Mechanosensory bristle	Mixed*
Wild type	177.5±8.6	114.5±3.5	70.3±1.5	0	0
amos^{1}	0	0	63.8±9.3	14.4±4.2	7.0±2.1
amos^{1}*/Df(2L)M36F-S6*	0	0	53.5±5.0	14.5±3.0	2.2±2.2
amos^{2}	9.6±5.6	0	62.8±7.1	13.3±4.5	15.2±3.8
amos^{2}*/Df(2L)M36F-S6*	5.8±3.4	0	52.0±1.4	12.5±3.9	21.5±2.6
amos^{3}*/Df(2L)M36F-S6*	0	0	72.6±6.4	20.8±0.8	3.6±2.2
lz^{34}; *amos*^{1}*/Df(2L)M36F-S6*	0	0	56.7±3.8	0	0
sc$^{10\text{-}1}$; *amos*^{1}	0	0	83.5±0.7	6.0±1.4	10.5±2.2
ase^{1}; *amos*^{1}*/Df(2L)M36F-S6*	0	0	58.2±8.3	14.8±2.2	6.8±2.9

Values are sensilla per antenna±s.d. (number of antenna scored, n=4-11).
'Mixed' refers to sensilla of undefinable morphology (*amos*^{1} and *amos*^{3}) or intermediate olfactory/bristle morphology (*amos*^{2}).

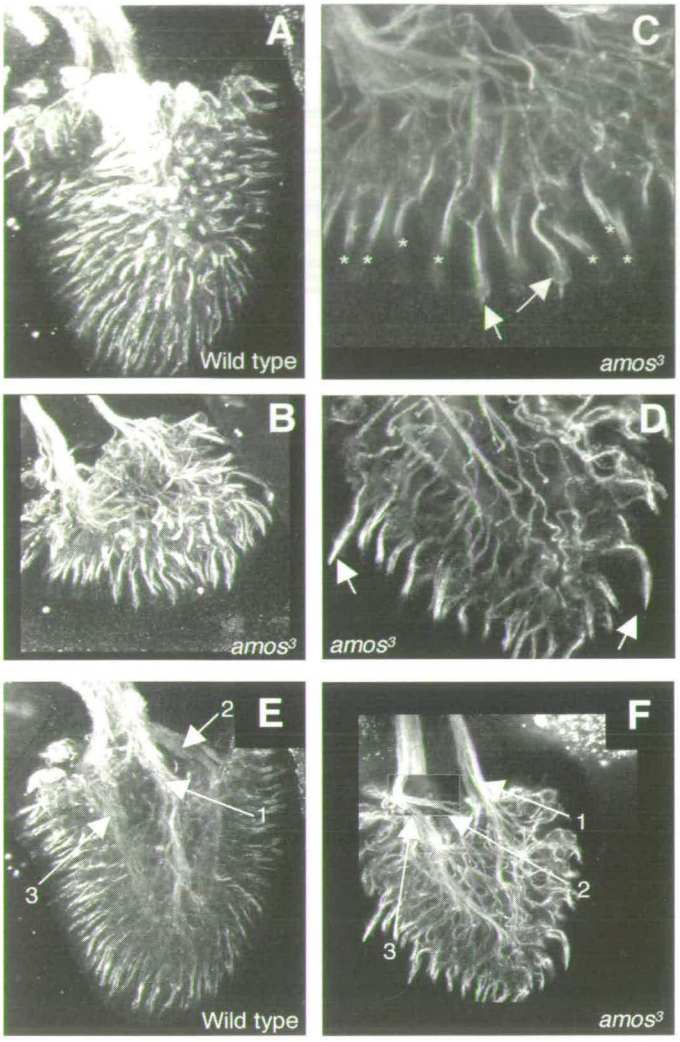


Fig. 2. Olfactory receptor neurons in *amos* mutants. Confocal projection images of late pupal antennae stained to detect sensory neurons. (A) Wild type. Clusters of cell bodies and their dendrites can be seen. (B) *amos*^{3} mutant showing far fewer clusters. (C,D) Higher magnification views. Although most ORNs are clustered, as seen by the multiple dendrites (*) (representing sensilla coeloconica), some sensilla appear to be mono-innervated (arrows) and may represent the bristles. (E) Wild-type confocal section showing the three olfactory nerve bundles. (F) Confocal section of *amos* mutant, with the three bundles labelled (a small section has been pasted in from another confocal plane to show clearly the second bundle).

present, although consisting of fewer axons as expected (comprising the axons of *ato*-dependent ORNs) (Fig. 2F). Although thinner, the fascicles appear normal in structure and location. Thus, in contrast to *ato* (Jhaveri et al., 2000b), mutations of *amos* do not cause defects in routing or fasciculation of the olfactory nerves. This supports the conclusion that the *ato*-dependent sensory lineage provides the information for fasciculation of these nerves.

amos expression prefigures a late, *ato*-independent subset of olfactory precursors

Olfactory precursors arise in the pupal antennal imaginal disc over an extended period of time (Ray and Rodrigues, 1995). Given their high density, the appearance of olfactory precursors is complex and incompletely characterised. We initially characterised the evolution of this pattern by studying Senseless (Sens; Lyra – FlyBase) expression, which is a faithful indicator of proneural-derived sensory precursors and is probably a direct target of proneural proteins (Nolo et al., 2000). We found that precursor formation occurs in three waves. First, Sens expression begins a few hours before puparium formation (BPF) in an outer semicircle of cells (Fig. 3A). A second wave begins at 0-4 hours after puparium formation (APF) to give a very characteristic pattern, including three semicircles of precursors (Fig. 3B). After this, a third wave appears over an extended period of time, with increasing numbers of cells appearing intercalated between the early precursors until no spatial pattern features can be observed (Fig. 3C,D).

Using a polyclonal antibody raised against the entire Amos protein, we determined that *amos* expression begins at puparium formation in three distinct semicircles and then continues for the next 16 hours, with the semicircles becoming indistinct by around 8 hours APF (Fig. 3E-H). The characteristic early waves of SOPs arise between the Amos domains of expression and do not show overlap with Amos expression (Fig. 3I-L). However, the third wave of SOPs appears to arise from the Amos expression domains. These late SOPs co-express Amos, and their nuclei lie beneath the Amos expression domains, consistent with these cells being olfactory SOPs (Fig. 4E,F and data not shown). Unusually, overlying *amos* proneural cluster expression is evidently not affected by lateral inhibition upon the appearance of the SOPs. *ato* is expressed much earlier than *amos*. All wave 1 and 2 SOPs appear to express Ato or to have arisen from Ato-expressing cells (see also Jhaveri et al., 2000a) (Fig. 4A-D). Consistent with this, the entire early SOP pattern is missing in antennal discs from *ato*^{1} mutant pupae (Fig. 5A-C). SOPs only begin to appear between 4 and 8 hours APF, corresponding to the third wave of precursors. These coincide very precisely with Amos expression, which itself appears unaffected (Fig. 5D-F).

In summary, there are three waves of olfactory precursor formation (Fig. 4G). The first and second waves are well defined, giving rise to the sensilla coeloconica of the sacculus and the antennal surface, respectively. These precursors express and require *ato*. The third wave of precursors is much more extensive and has little obvious pattern, giving rise to the more numerous sensilla basiconica and trichodea. Amos is expressed in a pattern entirely consistent with it being the proneural gene for the late wave precursors. Expression of Amos is complementary with that of Ato and is independent of Ato function. Thus, Ato- and Amos-expressing SOPs show a degree of spatial and temporal separation.

amos appears to be expressed in proneural domains and then in SOPs. For *sc* and *ato*, these two phases of expression are driven by separate enhancers, and SOP-specific enhancers have been identified (Culí and Modolell, 1998; Sun et al., 1998). A 3.6 kb fragment upstream from *amos* was found to support GFP reporter gene expression in the pupal third antennal segment. Comparison with Amos and Sens expression showed that GFP coincides with the Amos but not Ato SOPs (Fig. 6A,B). This fragment therefore contains an *amos* SOP enhancer. Perduring GFP expression driven by the enhancer

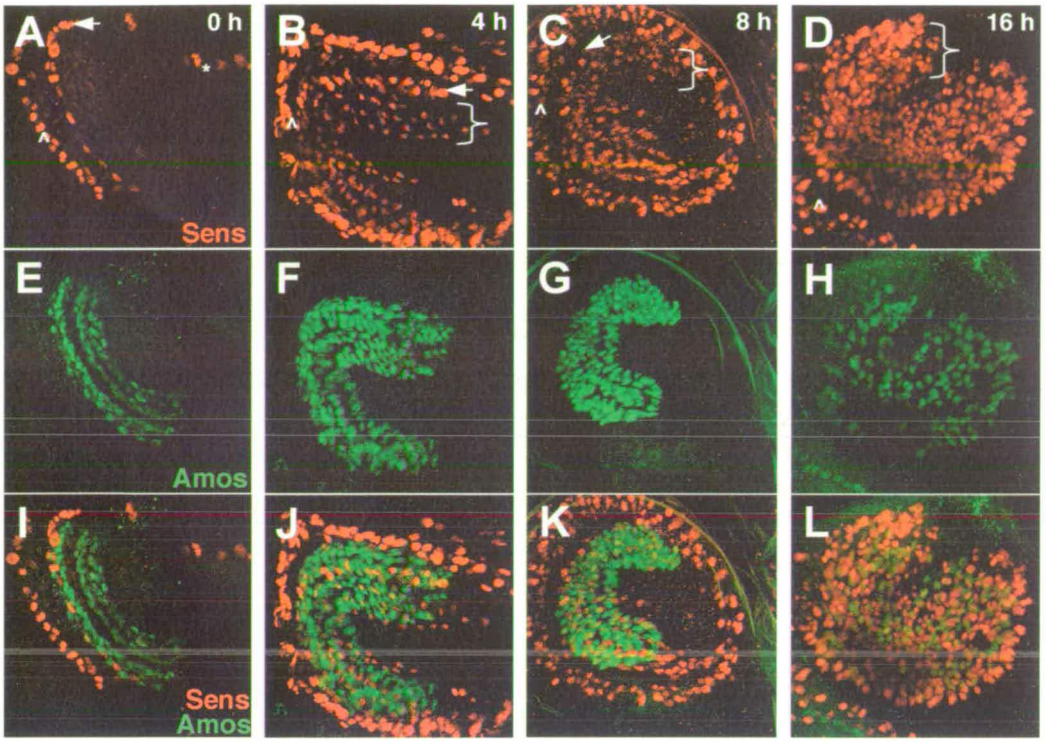


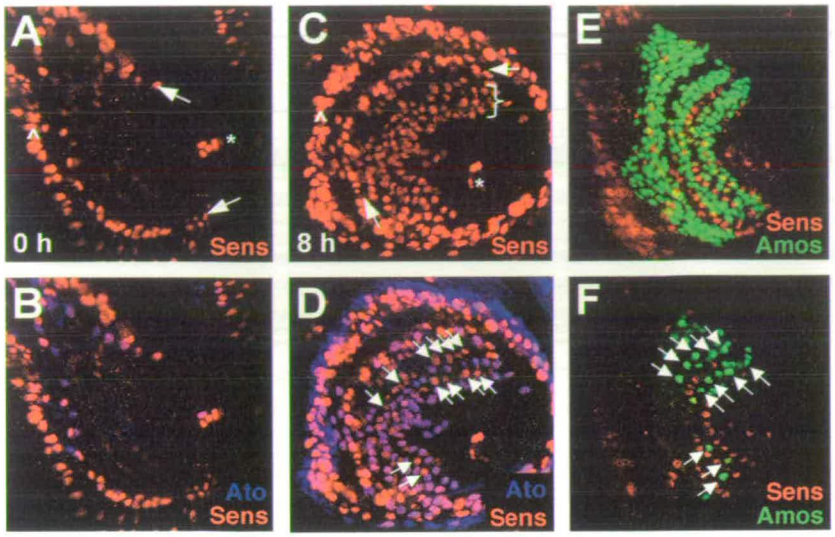
Fig. 3. Amos expression during olfactory precursor formation. (A-L) Time course of Amos protein expression relative to olfactory precursor formation. In all cases we concentrate on precursors in the third segment, although a large number of chordotonal precursors are also visible in the surrounding second segment (carets; see also the summary in Fig. 4). (A,E,I) At 0 hours APF, the first wave of precursors appear (arrow). (B,F,J) At 4 hours APF, the second wave of precursors appears in a highly characteristic pattern (bracket). (C,G,K) At 8 hours APF, the third wave of precursors accumulate between the rows of the second wave, eventually obscuring any clear pattern by 16 hours APF (D,H,L). (A-D) Amos expression is detected throughout this time, but the expression is ectodermal from 0-4 hours APF, and then it co-labels with some of the precursors between 4-16 hours APF. Amos continues to be expressed in some cells at 16 hours APF, and these cells seem to be a mixture of ectodermal cells and precursors.

can be observed in large numbers of sensilla on the maturing pupal antenna. From their morphology, it is clear that the GFP-expressing subset are the sensilla trichodea and basiconica (Fig. 6G). This confirms that early SOPs form sensilla coeloconica whereas late SOPs produce sensilla trichodea and basiconica. Interestingly, these GFP-expressing sensilla differentiate late, because there is no overlap with the 22C10 marker until late in development (Fig. 6C,D). Thus, the timing of neuronal differentiation reflects the timing of SOP birth. These findings correlate with the differing effects of proneural genes on fasciculation as described above: the first-born *ato*-dependent cells organise the nerves, and the later *amos*-dependent ORNs follow passively.

Loss of olfactory precursors in *amos* mutants

Loss of SOPs is one of the defining characteristics of proneural gene mutations. We examined SOP formation in *amos* mutants relative to wild type by examining Sens expression. As expected from the expression analysis described above, the first two waves of SOP formation show little discernible difference in pattern between *amos* mutant antennal discs and wild-type discs (Fig. 5G,H). This is consistent with these early SOPs expressing and requiring *ato*, and they indeed express *ato* in a pattern indistinguishable from wild type (Fig. 5K). This shows that *ato* expression does not depend on *amos* function. After this, the later arising SOPs do not appear to form between rows of *ato*-dependent SOPs, corresponding to those cells shown to express *amos* (compare Fig. 5K with Fig. 4D). A few precursors do not express *ato*, and these may represent precursors of the ectopic bristles (Fig. 5K). This is supported by an analysis of Cut expression, which is the key molecular switch that must be activated to allow SOPs to take a bristle fate (Bodmer et al., 1987; Blochlinger et al., 1991), and whose expression correlates with bristle SOPs (Blochlinger et al., 1990). In the wild-type antenna, Cut is not expressed during olfactory SOP formation (Fig. 5M), although later it is expressed in differentiating cells of all olfactory sensilla (Fig. 6I and data not shown). This expression normally appears after 16 hours and does not overlap with Amos. In *amos* mutant antennae, expression begins earlier than normal in a subset of SOPs that appear to correspond to the ones identified above (Fig. 5N).

By 16 hours APF, there is a large loss of Sens staining in *amos* mutants (Fig. 5I,L). The remaining cells tend to be in clusters as would be expected for the early *ato*-dependent sensilla, but otherwise the identity of these cells cannot be determined. Detection of Amos protein in $amos^{1}$ mutant antennal discs also shows that although the Amos domains are still present, the deeper Amos/Sens-expressing nuclei are absent (Fig. 5J). Thus, at least a large number of *amos*-associated SOPs are not formed in the *amos* mutant.



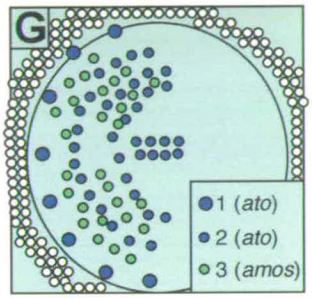


Fig. 4. Olfactory precursors and *amos/ato* expression. (A,B) *ato* is expressed in the wave 1 precursors. The arrows mark the ends of this semicircular line of olfactory precursors. *ato*-dependent arista precursors (*) and chordotonal precursors (caret) are also marked. (C,D) *ato* is expressed in the wave 2 olfactory precursors (bracket). It is not expressed in the third wave, the first cells of which can be seen between the Ato rows (arrows in D). (E,F) *amos* is expressed in a complementary way to *ato* in the third wave of precursors. (E) Confocal projection of a stack of images showing Amos detection in a similar disc to D. (F) Deep confocal section from E showing nuclei of *amos*-expressing precursors (arrows) underlying the main *amos* proneural expression domain. These correspond to the non-Ato-expressing precursors (arrows in D). (G) Schematic summary of *amos*- and *ato*-dependent precursor pattern at ~8 hours APF.

Expression of *amos* during sensillum development

The processes and lineages by which olfactory SOPs lead to the differentiated cells of the olfactory sensillum are not entirely known. The limited information available comes from analysis of the early wave of SOPs, which we have established represent the *ato*-dependent sensilla. After an SOP is selected there appears in its place a cluster of 2-3 cells expressing the A101 enhancer trap [the pre-sensillum cluster (PSC)]; this is apparently caused not by division of the SOP but perhaps by recruitment by the SOP (Ray and Rodrigues, 1995; Reddy et al., 1997), although the evidence for this is indirect. These PSC cells then divide to form the cells of the sensillum, including the outer support cells (hair and socket cells), inner support cells (sheath cells) and 1-4 neurons. For the early subset of SOPs, formation of the PSC occurs at a time in which *amos* is still expressed in the epithelial domains, and so *amos* could influence the development of these cells. Using A101 as a marker of the PSC cells, we determined that *amos* is not expressed in recognisable PSCs at 8 or 16 hours APF (Fig. 6E). Moreover, there is also no apparent co-labelling of Amos and Pros [a marker of one of the PSC cells (Sen et al., 2003)] (Fig. 6F). This suggests either that early PSC cells do not derive from *amos*-expressing cells or that *amos* is switched off rapidly when cells join a PSC.

The situation appears different for the cells derived from *amos*-dependent SOPs. Surprisingly at 24 hours and beyond, the *amos* enhancer drives GFP expression in most or all cells of the differentiating sensilla basiconica and trichodea (Fig. 6G): including most or all of the neurons (recognised by Elav expression; Fig. 6H); the sheath cell (recognised by Pros expression; Fig. 6H); and the outer support cells (recognised by the higher expression of Cut; Fig. 6I). This suggests that the late PSC cells do derive from *amos*-expressing cells and that activation of an enhancer within the 3.6 kb regulatory fragment (possibly separate from the SOP enhancer) is part of their specification process, although *amos* expression itself may not be long lived in these cells.

amos represses *scute* function

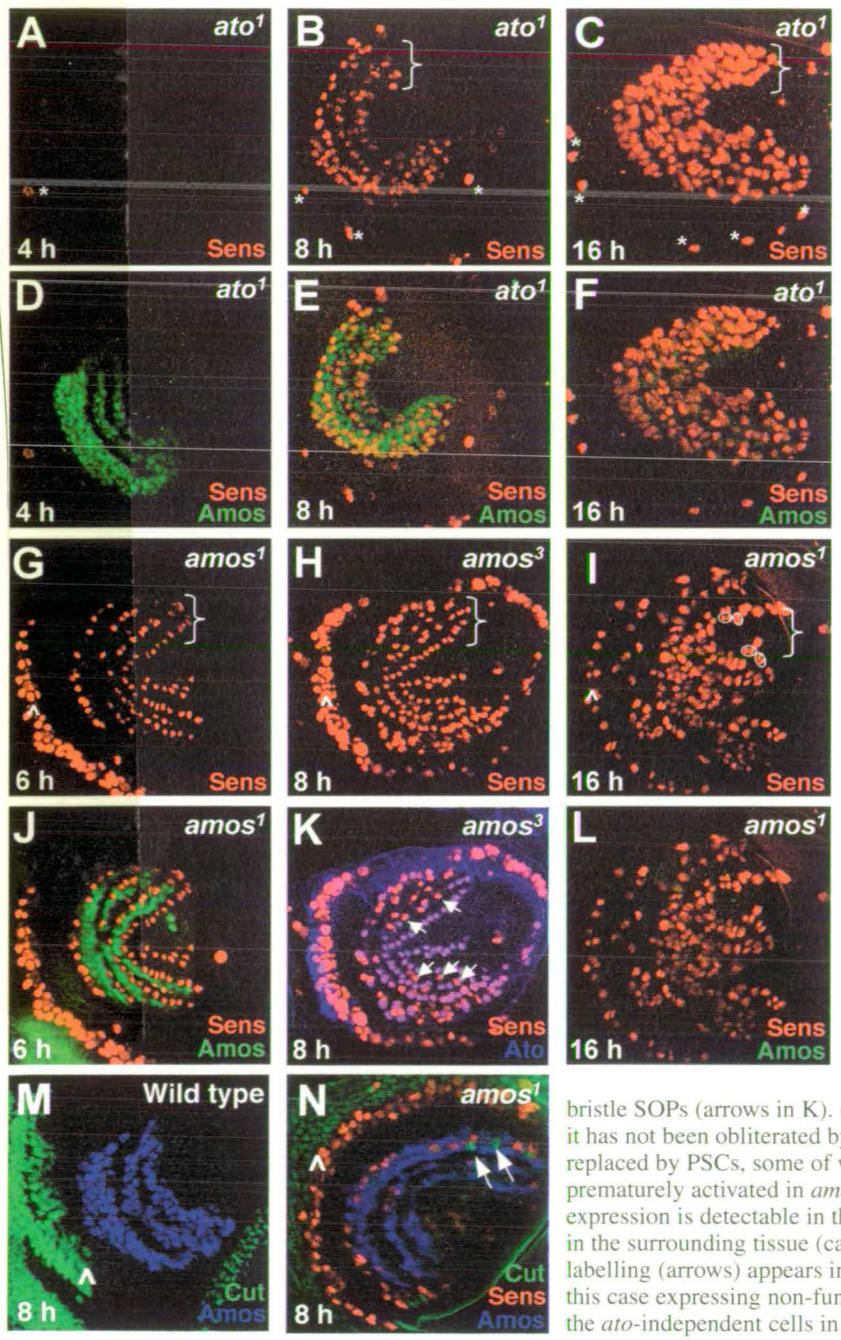
amos mutant antennae have Cut-expressing SOPs, but, although *cut* expression decides SOP subtype fate, it does not specify ectodermal cells as SOPs de novo. To investigate the involvement of other proneural genes, we first determined whether the bristles depended on *ato*, as it is expressed in close proximity to the emerging bristle SOPs. Clones of $amos^1$ mutant tissue were induced in ato^1 mutant antennae. In such clones, all olfactory sensilla were absent, as expected, but ectopic bristles were still formed (Fig. 1F). Therefore the bristles do not depend on *ato* function.

Cut expression normally follows from *ac/sc* proneural function, and so the ectopic bristle SOPs might depend on these proneural genes. Indeed, mutation of *ac* and *sc* greatly reduces the number of ectopic bristles in $amos^1$ flies ($In(1)sc^{10\text{-}1}/Y$; $amos^1/Df(2L)M36F\text{-}S6$ flies) (Fig. 1G and Table 2). By contrast, mutation of the non-proneural *ASC* gene *asense* (*ase*) had no effect alone (Table 2). This suggests that in the absence of *amos*, *ac/sc* function, to a large extent, causes the formation of bristle SOPs.

To determine how *amos* might normally repress bristle formation, we examined the pattern of *sc* mRNA in the pupal antenna. Significantly, a weak stripe of *sc* expression was observed in the wild-type antenna. (Fig. 7A). This stripe coincides with *amos* expression, and consists of ectodermal cells and SOPs (Fig. 7B,C). In the *amos* mutant antenna, *sc* mRNA expression was stronger and more clearly correlated with SOPs (Fig. 7C). This suggests that *sc* is expressed in olfactory regions of the wild-type antenna but that its function is repressed by the presence of *amos*. We therefore investigated *sc* functional activity in the antenna by analysing the expression of specific *sc* target genes as indicators of Sc protein function. Firstly, we examined Ac protein, whose expression is ordinarily activated by Sc function as a result of cross regulation (Gomez-Skarmeta et al., 1995). Ac protein is present in some SOPs in *amos* mutant antennae, but is not present in wild-type antennae (Fig. 7E,F). A similar result was observed for *sc-SOP-GFP*, which is a reporter gene construct that is directly activated by *sc* upon SOP formation (L. Powell and A.P.J., unpublished) (Culí and Modolell, 1998). This reporter showed GFP expression in some SOPs in *amos* mutant antennae but not in

wild-type antennae (data not shown). Finally, we examined *sc-E1-GFP*, a reporter gene construct comprising GFP driven solely by a *sc*-selective DNA binding site (L. Powell and A.P.J., unpublished) (Culí and Modolell, 1998). This reporter is invariably activated in all cells containing active Sc protein (including PNCs and SOPs) (L. Powell and A.P.J., unpublished). As with the other target genes, this reporter was only expressed in *amos* mutant antennae (Fig. 7G,H). Thus, we conclude that *sc* mRNA is expressed in the wild-type pupal antenna, and *amos* normally must repress either the translation of this RNA or the function of the Sc protein produced. This conclusion is supported by misexpression experiments. When *amos* is misexpressed in *sc* PNCs of the wing imaginal disc (*109-68Gal4*/UAS-*amos*) there is a dramatic reduction in bristle formation (Fig. 8A,B), even though endogenous *sc* RNA levels are unaffected (data not shown).

The transcription factor encoded by *lozenge* (*lz*) plays a number of roles in olfactory sensillum development, including activating *amos* expression (Goulding et al., 2000). Mutants therefore show a loss of many *amos*-dependent sensilla. Interestingly flies mutant for both *lz* and *amos* (lz^{34}; $amos^{1}$/*Df(2L)M36F-S6*) have third antennal segments that bear only sensilla coeloconica, and so the ectopic bristles of *amos* mutants are dependent on *lz* function (Table 2, Fig. 1H). Correlating with this, the expression of *sc* mRNA in the third antennal segment was much reduced in a *lz* mutant compared with wild type (Fig. 7D). Thus, *lz* appears at least partly responsible for the expression of *sc* in the antenna.



Discussion

We show definitively that *amos* is the proneural gene for the precursors of two classes of olfactory sensilla. These precursors are absent in *amos* mutants, resulting in highly defective antennae lacking all sensilla basiconica and trichodea. Unusually, this is not the only phenotype of *amos* mutants. Unique among *Drosophila* proneural genes, mutation of *amos* results in the appearance of new sense organs: mechanosensory bristles are now formed on the third antennal segment. We provide evidence that *amos* must normally repress *sc*-promoted bristle specification in addition to promoting olfactory neurogenesis. Significantly, inhibitory interactions between bHLH genes have recently been reported during mouse neurogenesis, where discrete domains of bHLH transcription factor expression are set up partly by mutual cross-inhibition combined with autoregulation (Gowan et

Fig. 5. Olfactory precursors in *amos* and *ato* mutants. These discs should be compared with the corresponding wild-type discs in Figs 3 and 4. (A,D) The early precursors are specifically lost in *ato* mutants. The remaining olfactory precursors correspond to the third wave (B,C) and align very closely with the *amos* expression domains (E,F). In the second segment, the chordotonal precursors are also missing and only a few bristle precursors remain (*, A-C). (G-L) The late precursors are specifically lost in *amos* mutants. (G,J) Early precursor pattern resembles wild type, with mutant $Amos^{1}$ protein detectable between the rows of precursors (brackets). Caret in G indicates chordotonal precursors. (H,K) At 8 hours APF, the pattern remains unchanged as the third wave SOPs are not formed (c.f. Fig. 4C). These early precursors mostly express Ato, although a number of non-Ato expressing SOPs appear between the early rows, which could correspond to the bristle SOPs (arrows in K). (I,L) The early pattern is still apparent at 16 hours APF as it has not been obliterated by the third wave of SOPs (the early SOPs have now been replaced by PSCs, some of which are ringed). (M,N) Cut expression appears prematurely activated in *amos* mutants. (M) Wild type at 8 hours APF. No Cut expression is detectable in the third segment SOPs; however, Cut stains very strongly in the surrounding tissue (caret). (N) $amos^{1}$ mutant at 8 hours APF. Some Cut labelling (arrows) appears in SOPs derived from the Amos-expressing domains (in this case expressing non-functional $Amos^{1}$ protein). These cells seem to correspond to the *ato*-independent cells in K.

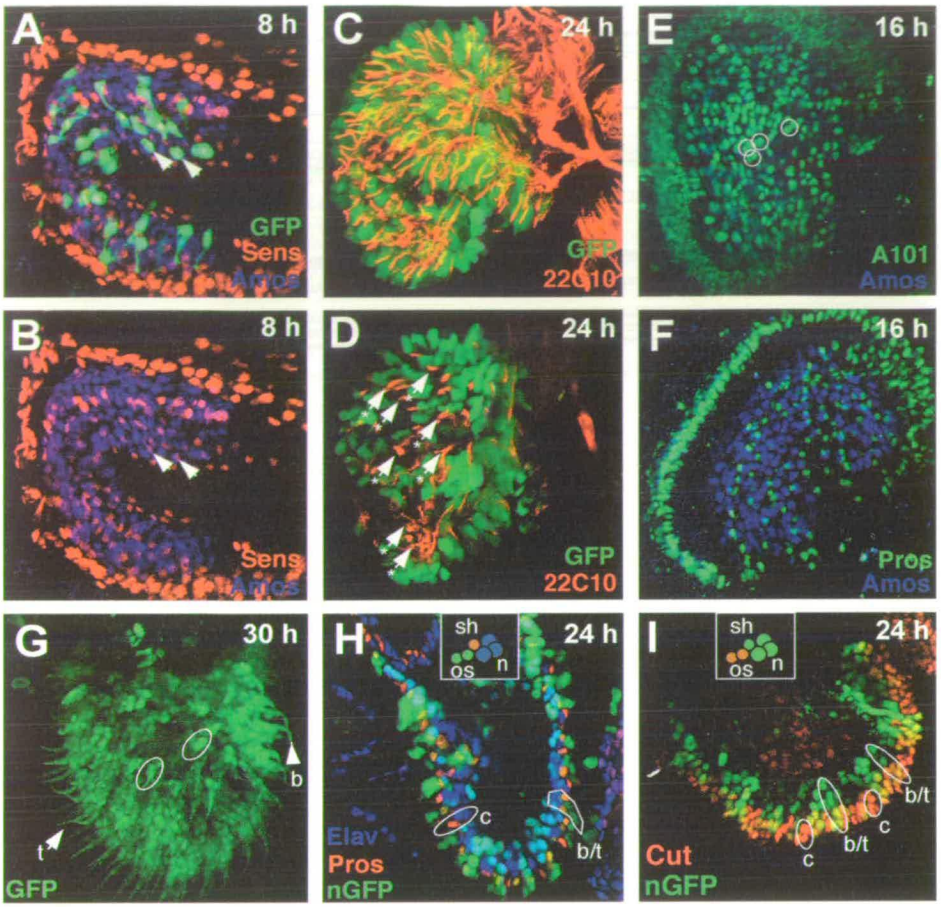


Fig. 6. Fate of *amos*-expressing cells during olfactory development. (A-D,G) Activity of an *amos* SOP enhancer driving GFP expression. (A,B) At 8 hours APF, GFP can be detected in the third wave of olfactory precursors, some co-labelled SOPs are indicated by arrows (co-labelled with Sens and Amos in the separation in B). (C,D) The *amos-GFP* expressing cells contribute to late differentiating sensilla, as shown by lack of co-labelling with a neuronal marker (22C10) at 24 hours APF. C is a projection of many sections whereas D is a confocal section with some of the differentiating neurons marked by asterisks, these do not express GFP. (E,F) Later expression of *amos* does not correspond to PSC cells. (E) At 16 hours APF, Amos expression is fading, but there is no overlap with A101 β-galactosidase expression in the PSCs, some of which are ringed. (F) There is no overlap of Amos expression with that of Pros, a marker of one of the PSCs. (G) *amos-GFP* construct at 30 hours APF: a large number of sensilla retain GFP. Protein appears to be in sensillar groups (as indicated by rings), and includes the outer support cells, so that sensilla trichodea (t) and basiconica (b) can clearly be discerned. (H,I) Analysis of *amos-GFP* in confocal sections of antennae at 24 hour APF relative to the component cells of the sensilla (see insets). n, neuron; sh, sheath; os, outer support cells. *amos-GFP* labels rows of cells corresponding to each sensillum basiconicum or trichodeum (some are ringed), whereas presumptive coeloconica (c) do not express GFP. (H) GFP is expressed in neurons (marked by Elav) and sheath cells (marked by Pros). (I) GFP is expressed in outer support cells (marked by stronger expression of Cut).

al., 2001; Nieto et al., 2001). As with *amos*, cross-inhibition occurs between members of different bHLH families: *Mash1* (ASC homologue), *Math1* (*ato* homologue), and *neurogenin1* (*tap* homologue).

How proneural genes determine neuronal subtype

On misexpression evidence, we have argued that neuronal subtype specification involves repression of bristle fate by *ato* during chordotonal SOP formation (Jarman and Ahmed, 1998) and by *amos* during olfactory precursor formation (Goulding et al., 2000). In this light, the ectopic bristles in *amos* mutants are of significant interest. They represent the first loss-of-function evidence that an *ato*-type proneural gene suppresses bristle fate during the normal course of its function. However, how this relates to *amos* function is complex. In misexpression experiments, bristle suppression by *amos* is most strongly observed using a PNC- and SOP-specific Gal4 driver line (Goulding et al., 2000) (this report). Yet paradoxically, misexpression of *amos* more generally in the ectoderm, but only weakly in SOPs, yields dramatically different results: in such cases *amos* produces ectopic bristles very efficiently (Huang et al., 2000; Lai, 2003; Villa Cuesta et al., 2003). This bristle formation does not require the function of endogenous *ac/sc* genes (Lai, 2003), but probably reflects the intrinsic SOP-specifying function of *amos* in situations that are not conducive to its subtype-specifying (and bristle suppressing) function. It appears therefore that bristle suppression particularly requires *amos* expression in SOPs.

What does *amos* repress in the antenna? It appears that *sc* is expressed within the wild-type *amos* expression domain during olfactory SOP formation. Clearly *amos* must prevent the function of *sc*, as *sc* expression in ectoderm usually results in bristle specification. It may be significant that some of the *sc* RNA is in olfactory SOPs in the wild-type antenna, suggesting that the SOP may be a major location of repression by *amos*, as indicated by misexpression experiments. However, some bristle formation is maintained in *ac/sc; amos* mutants. This may be due to redundancy with other genes in the ASC: certainly wild-type bristle formation outside the antenna is not completely abolished in the absence of *ac/sc* (A.P.J., unpublished). An alternative possibility is that some bristle SOPs result from other proneural-like activity in the antenna. Direct proneural activity of *lz* is a possibility, although misexpression of *lz* elsewhere in the fly (using a *hs-lz* construct) is not sufficient to promote bristle formation (P.I.z.L., unpublished).

The *amos*^{2} hypomorph appears to represent a different situation. In such flies, a number of *amos*-dependent SOPs appear to have mixed olfactory/bristle fate. This suggests that on occasions the mutant Amos^2 protein is able to specify SOPs, but is less able to impose its subtype function (and so this, to some extent, resembles more the outcome of some misexpression experiments). *amos*^{2} may therefore be a useful

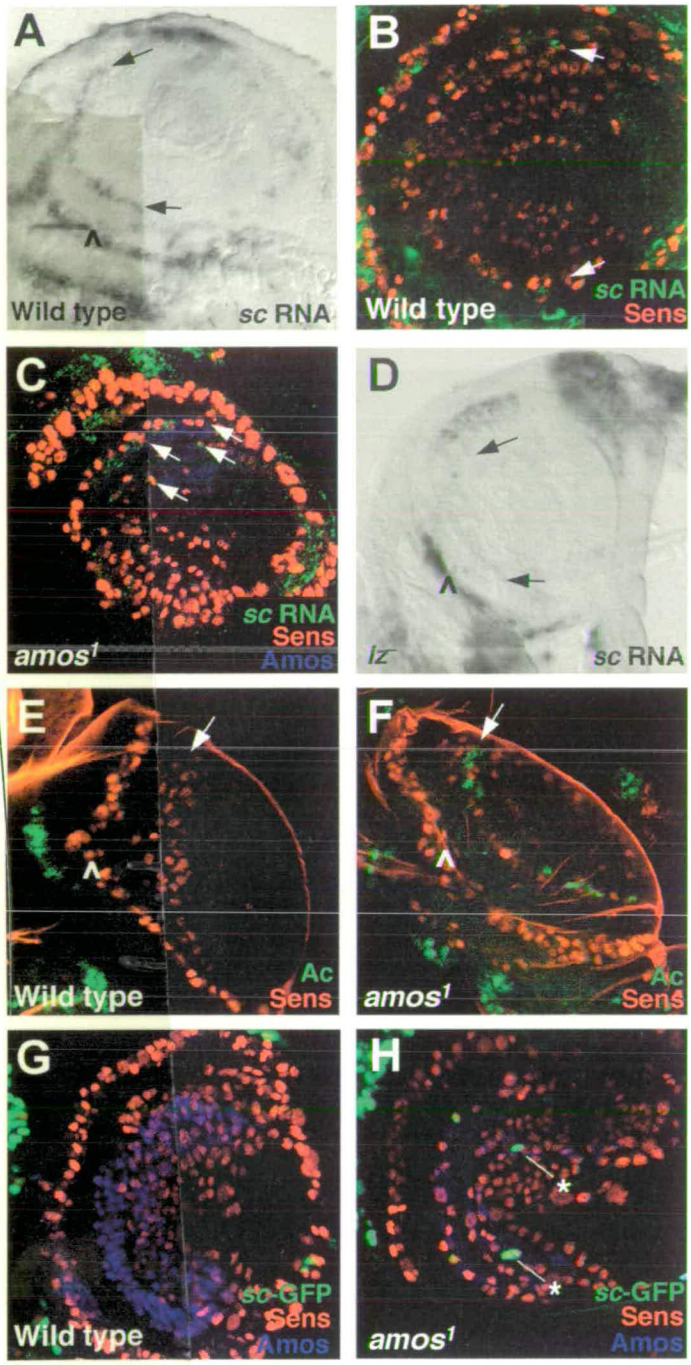


Fig. 7. Expression of *sc* and *sc* target genes in the antenna. (A-D) *sc* mRNA detected by in situ hybridisation. (A) Wild type, with *sc* expressed not only in the second antennal segment (caret) but also in the third segment (arrows). (B) Wild type, with sc RNA detected by immunofluorescence (green). (C) $amos^1$ mutant. *sc* mRNA is increased and is present in SOPs (arrows). (D) The second segment *sc* expression is reduced in lz^{34} mutants. (E,F) Ac expression is present in some SOPs in *amos* mutants. (E) Wild type at 8 hours APF, showing very little Ac expression in the third segment (first precursor wave marked by arrow) (some is visible in the second segment; caret). (F) $amos^3$ mutant at 8 hours APF, showing some Ac expression in second segment (arrow). (G,H) GFP expression from *sc-E1-GFP* reporter transgene. (G) Wild type, showing no expression in third segment. (H) $amos^3$ mutant showing expression (*).

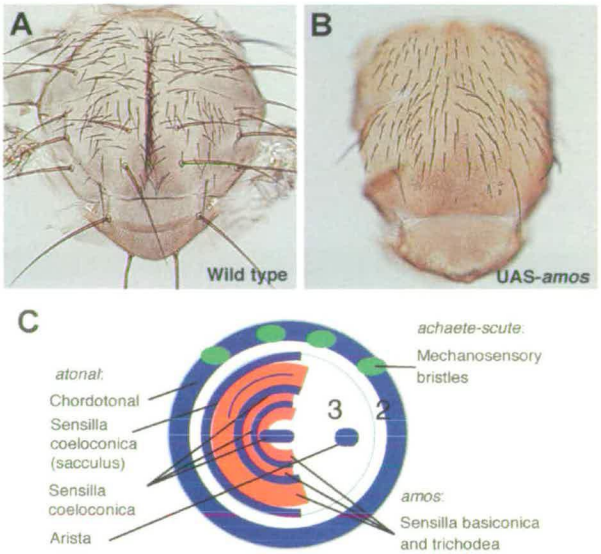


Fig. 8. *amos* misexpression represses bristle formation. (A) Wild type dorsal thorax. (B) Dorsal thorax from *109-68Gal4*/UAS-*amos* fly. *amos* misexpression driven in *sc* PNCs by this driver line results in loss of many bristles (mainly the large macrochaetae). (C) Summary of proneural functions in antenna. Diversity of sense organs laid down by function of three proneural gene systems. Blue, *atonal*; red, *amos*; green, *achaete/scute*.

tool for exploring these two functions. For example, if subtype specification requires interaction of Amos with protein co-factors (Jarman and Ahmed, 1998; Brunet and Ghysen, 1999; Hassan and Bellen, 2000), then these interactions may be specifically impaired in the $amos^2$ mutant.

Because the proneural proteins are normally transcriptional activators, it is unlikely that Amos/Ato proteins directly inhibit gene expression during bristle suppression (Jarman and Ahmed, 1998). The presence of *sc* RNA in *amos*-expressing cells in the wild-type antenna is consistent with this. The involvement of protein interactions is to be suspected. An interesting parallel is found in vertebrates, where neurogenin1 promotes neurogenesis and inhibits astrocyte differentiation (Nieto et al., 2001). The glial inhibitory effect could be separated from the neurogenesis promoting effect: whereas neurogenesis promotion depends on DNA binding and activation of downstream target genes, astrocyte differentiation was inhibited through a DNA-independent protein-protein interaction with CBP/p300 (Sun et al., 2001; Vetter, 2001). In the case of *amos*, an interesting possibility is that inhibition of bristle formation may involve the sequestering of Sc protein by Amos protein. As discussed above, such a mechanism would have to be sensitive to the level or pattern of *amos*, as general misexpression does not mimic this activity.

Comparison of *amos* and *ato* as olfactory proneural genes

Apart from giving rise to separate classes of olfactory precursor, there are interesting differences in the way that *ato* and *amos* are deployed in the antenna. We characterised three waves of olfactory precursor formation (Fig. 4G). The first and second waves are well defined, giving rise to well-patterned

sensilla coeloconica of the sacculus and the antennal surface, respectively. These precursors express and require *ato*. The third wave of precursors is much more extensive and has little obvious pattern; it gives rise to the much more numerous sensilla basiconica and trichodea. This wave expresses and requires *amos*. For the early waves, *ato* is expressed according to the established paradigm: it is expressed in small PNCs, each cluster giving rise to an individual precursor (Gupta and Rodrigues, 1997). The pattern of the PNCs is very precise and prefigures the characteristic pattern of precursors. *amos* expression is dramatically different. It is expressed in large ectodermal domains for an extended period of time. Densely packed precursors arise from this domain continuously without affecting the domain expression. This shows that singling out does not necessarily require shut down of proneural expression, and therefore has implications for how singling out occurs. In current models, it is assumed that PNC expression must be shut down to allow an SOP to assume its fate. The *amos* pattern better supports the idea that a mechanism of escaping from or becoming immune to lateral inhibition is more likely to be important generally. One prediction would be that *amos* and *ato* (and *ac/sc*) differ in their sensitivities to Notch-mediated lateral inhibition, a situation that has been noted for mammalian homologues (Lo et al., 2002).

Why are the proneural genes deployed so differently? One possibility is simply that there are very many more sensilla basiconica and trichodea than coeloconica. All the coeloconica precursors can be formed by *ato* action in a precise pattern in two defined waves. This would not be possible for the large number of basiconica and trichodea precursors, and so precursor selection has been modified for *amos*. Indeed, *amos* appears to be a particularly 'powerful' proneural gene when misexpressed (Lai, 2003; Villa Cuesta et al., 2003). This may make *amos* a useful model of other neural systems in which large numbers of precursors must also be selected.

For most insects, the antenna is the major organ of sensory input. It is not only the site of olfaction, but also of thermoreception, hygroreception, vibration detection and proprioception, as well as of touch. Patterning the sensilla is therefore complex and three types of proneural gene are heavily involved to give different SOPs (Fig. 7G). It is clear that the study of antennal sensilla will provide a useful model for exploring the fate determining contribution of intrinsic bHLH protein specificity and extrinsic competence factors.

We wish to thank Anindya Sen for help and discussions. We thank Hugo Bellen for the anti-Sens antibodies, and the Developmental Biology Hybridoma Bank, Iowa for other antibodies. This work was supported by a Wellcome Trust project grant (055851) and Senior Research Fellowship to A.P.J. (042182).

References

Bertrand, N., Castro, D. S. and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517-530.

Blochlinger, K., Bodmer, R., Jan, L. Y. and Jan, Y. N. (1990). Patterns of expression of Cut, a protein required for external sensory organ development in wild-type and *cut* mutant *Drosophila* embryos. *Genes Dev.* **4**, 1322-1331.

Blochlinger, K., Jan, L. Y. and Jan, Y. N. (1991). Transformation of sensory organ identity by ectopic expression of Cut in *Drosophila*. *Genes Dev.* **5**, 1124-1135.

Bodmer, R., Barbel, S., Shepherd, S., Jack, J. W., Jan, L. Y. and Jan, Y. N. (1987). Transformation of sensory organs by mutations of the *cut* locus of *D. melanogaster*. *Cell* **51**, 293-307.

Brunet, J.-F. and Ghysen, A. (1999). Deconstructing cell determination: proneural genes and neuronal identity. *BioEssays* **21**, 313-318.

Carlson, J. R. (1996). Olfaction in *Drosophila*: from odor to behavior. *Trends Genet.* **12**, 175-180.

Culí, J. and Modolell, J. (1998). Proneural gene self-stimulation in neural precursors: an essential mechanism for sense organ development that is regulated by *Notch* signalling. *Genes Dev.* **12**, 2036-2047.

Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., FerresMarco, D., Beamonte, D. and Modolell, J. (1995). Cis-regulation of *achaete* and *scute*: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes Dev.* **9**, 1869-1882.

Goulding, S. E., zur Lage, P. and Jarman, A. P. (2000). *amos*, a proneural gene for *Drosophila* olfactory sense organs that is regulated by *lozenge*. *Neuron* **25**, 69-78.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R. and Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron* **31**, 219-232.

Gupta, B. P. and Rodrigues, V. (1997). *atonal* is a proneural gene for a subset of olfactory sense organs in *Drosophila*. *Genes Cells* **2**, 225-233.

Hassan, B. A. and Bellen, H. J. (2000). Doing the MATH: is the mouse a good model for fly development? *Genes Dev.* **14**, 1852-1865.

Hassan, B. A., Bermingham, N. A., He, Y., Sun, Y., Jan, Y. N., Zoghbi, H. Y. and Bellen, H. J. (2000). *atonal* regulates neurite arborization but does not act as a proneural gene in the *Drosophila* brain. *Neuron* **25**, 549-561.

Huang, M. L., Hsu, C. H. and Chien, C. T. (2000). The proneural gene *amos* promotes multiple dendritic neuron formation in the *Drosophila* peripheral nervous system. *Neuron* **25**, 57-67.

Jarman, A. P. and Ahmed, I. (1998). The specificity of proneural genes in determining *Drosophila* sense organ identity. *Mech. Dev.* **76**, 117-125.

Jarman, A. P., Grau, Y., Jan, L. Y. and Jan, Y. N. (1993). *atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. *Cell* **73**, 1307-1321.

Jarman, A. P., Grell, E. H., Ackerman, L., Jan, L. Y. and Jan, Y. N. (1994). *atonal* is the proneural gene for *Drosophila* photoreceptors. *Nature* **369**, 398-400.

Jhaveri, D., Sen, A., Reddy, G. V. and Rodrigues, V. (2000a). Sense organ identity in the *Drosophila* antenna is specified by the expression of the proneural gene *atonal*. *Mech. Dev.* **99**, 101-111.

Jhaveri, D., Sen, A. and Rodrigues, V. (2000b). Mechanisms underlying olfactory neuronal connectivity in *Drosophila* – the *atonal* lineage organizes the periphery while sensory neurons and glia pattern the olfactory lobe. *Dev. Biol.* **226**, 73-87.

Johnston, L. A., Ostrow. B. D., Jasoni, C. and Blochlinger, K. (1998). The homeobox gene *cut* interacts genetically with the homeotic genes *proboscipedia* and *Antennapedia*. *Genetics* **149**, 131-142.

Lai, E. C. (2003). *Drosophila Tufted* is a gain-of-function allele of the proneural gene *amos*. *Genetics* **163**, 1413-1425.

Lindsley, D. L. and Zimm, G. G. (1992). *The Genome of* Drosophila melanogaster. San Diego: Academic Press.

Lo, L., Dormand, E., Greenwood, A. and Anderson, D. J. (2002). Comparison of generic neuronal differentiation and neuron subtype specification functions of mammalian *achaete-scute* and *atonal* homologues in cultured neural progenitor cells. *Development* **129**, 1553-1567.

Newsome, T. P., Asling, B. and Dickson, B. J. (2000). Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851-860.

Nieto, M., Schuurmans, C., Britz, O. and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. *Neuron* **29**, 401-413.

Nolo, R., Abbott, L. A. and Bellen, H. J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102**, 349-362.

Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J. and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. *Genes Dev.* **16**, 324-338.

Ray, K. and Rodrigues, V. (1995). Cellular events during development of the olfactory sense organs in *Drosophila melanogaster*. *Dev. Biol.* **167**, 426-438.

Reddy, G. V., Gupta, B., Ray, K. and Rodrigues, V. (1997). Development of the *Drosophila* olfactory sense organs utilizes cell-cell interactions as well as lineage. *Development* **124**, 703-712.

Sen, A., Reddy, G. V. and Rodrigues, V. (2003). Combinatorial expression of Prospero, Seven-up and Elav identifies progenitor cell types during sense-organ differentiation in the *Drosophila* antenna. *Dev. Biol.* **254**, 79-92.

Shanbhag, S. R., Müller, B. and Steinbrecht, R. A. (1999). Atlas of olfactory organs of *Drosophila melanogaster*. 1. Types, external organization, innervation and distribution of olfactory sensilla. *Int. J. Insect Morph. Embryol.* **28**, 377-397.

Sun, Y., Jan, L. Y. and Jan, Y. N. (1998). Transcriptional regulation of *atonal* during development of the *Drosophila* peripheral nervous system. *Development* **125**, 3731-3740.

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G. and Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365-376.

Vetter, M. (2001). A turn of the helix: preventing the glial fate. *Neuron* **29**, 559-562.

Villa Cuesta, E., de Navascues, J., Ruiz Gomez, M., del Corral, R. D., Dominguez, M., de Celis, J. F. and Modolell, J. (2003). *Tufted* is a gain-of-function allele that promotes ectopic expression of the proneural gene *amos* in *Drosophila*. *Genetics* **163**, 1403-1412.