Chem. Senses 26: 201–206, 2001

Olfaction in *Drosophila*: Coding, Genetics and e-Genetics

Coral Warr[1], Peter Clyne[1], Marien de Bruyne[1], Junhyong Kim[1,2] and John R. Carlson[1]

[1]Department of Molecular, Cellular, and Developmental Biology and [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520-8103, USA

Correspondence to be sent to: John Carlson, Department of Molecular, Cellular, and Developmental Biology, Yale University, PO Box 208103, New Haven CT 06520-8103, USA. e-mail john.carlson@yale.edu

Abstract

Odor coding in *Drosophila* is examined at both the cellular and molecular levels. Functional analysis of individual olfactory receptor neurons (ORNs) by single-unit electrophysiology has shown that ORNs divide into discrete classes, with each class exhibiting a characteristic odor response spectrum. Extensive analysis of ORNs in the maxillary palp has revealed six such classes, which are combined in sensilla according to a strict pairing rule. In order to identify the odor receptor genes that determine the odor specificity of these ORN classes, a new algorithm was designed to search DNA databases for proteins with a particular structure, as opposed to a particular sequence. The algorithm identified a large family of genes likely to encode odor receptors. The *acj6* gene, originally identified in a screen for mutants defective in olfactory behavior, encodes a transcription factor that regulates a subset of these receptor genes, and is likely to play a critical role in the process by which ORNs select which receptors to express.

Introduction

Drosophila is an attractive organism in which to study olfaction, in several respects. First, its olfactory system is relatively simple, containing only ~10^3 receptor neurons (Stocker, 1994), whereas humans have 10^8. Second, there are powerful genetic approaches that can be used in *Drosophila* to identify and characterize genes required for olfactory system function and development. Third, the *Drosophila* genome has been sequenced (Adams *et al.*, 2000; Rubin *et al.*, 2000). Finally, and perhaps most compelling, olfactory function can be conveniently analyzed *in vivo*, by measuring either the electrophysiological or behavioral response (Siddiqi, 1991; Carlson, 1996).

Identification of olfactory receptor neuron classes by single-unit electrophysiology

In order to investigate the cellular basis of olfactory coding in *Drosophila*, we have analyzed the function of individual olfactory receptor neurons (ORNs) directly, by single-unit electrophysiology (Schneider and Boeckh, 1962; Boeckh and Ernst, 1987). Two central issues of interest are the number of different types of ORNs and the breadth of their specificities. A priori, there could be 1000 different neuronal types, each type sensitive to a single, unique odor. Another possibility is that there could be a much smaller number of types, each sensitive to a broad spectrum of odors.

As a first step in the functional analysis of the cellular basis of coding, we examined the maxillary palp (de Bruyne *et al.*, 1999). This organ contains only 120 ORNs, compartmentalized in sensilla. We have measured the response of individual neurons in these sensilla to a wide variety of odors. These measurements are made by placing an electrode into the base of a sensillum and then recording the activity of the neurons following odor stimulation (Kaissling, 1995). In general one can distinguish the responses of the different neurons in a sensillum because they produce action potentials that differ in height and/or shape. An example of such a recording, from a sensillum stimulated with ethyl acetate, can be seen in Figure 1a. A burst of action potentials is observed, and inspection reveals that only one of the two neurons responds, the one with the large spikes, which we call the A neuron. The B neuron, with the smaller spikes (indicated by dots in the figure), does not respond to ethyl acetate.

How broad are the response spectra of these ORNs? The spectra of two kinds of neuron are shown in Figure 1b. One of them responds most strongly to ethyl acetate, among the 16 chemically diverse stimuli shown, but it also responds less strongly to several other odors. The other appears to be narrowly tuned. It responds strongly to 4-methylphenol, but the only other tested odor that elicits even a weak response is 4-methylcyclohexanol, whose structure is related to that of 4-methylphenol.

We have defined six functional classes of ORNs, on the basis of hundreds of recordings from maxillary palp sensilla. Each sensillum contains two physiologically distinct neurons, which is fully consistent with anatomical data

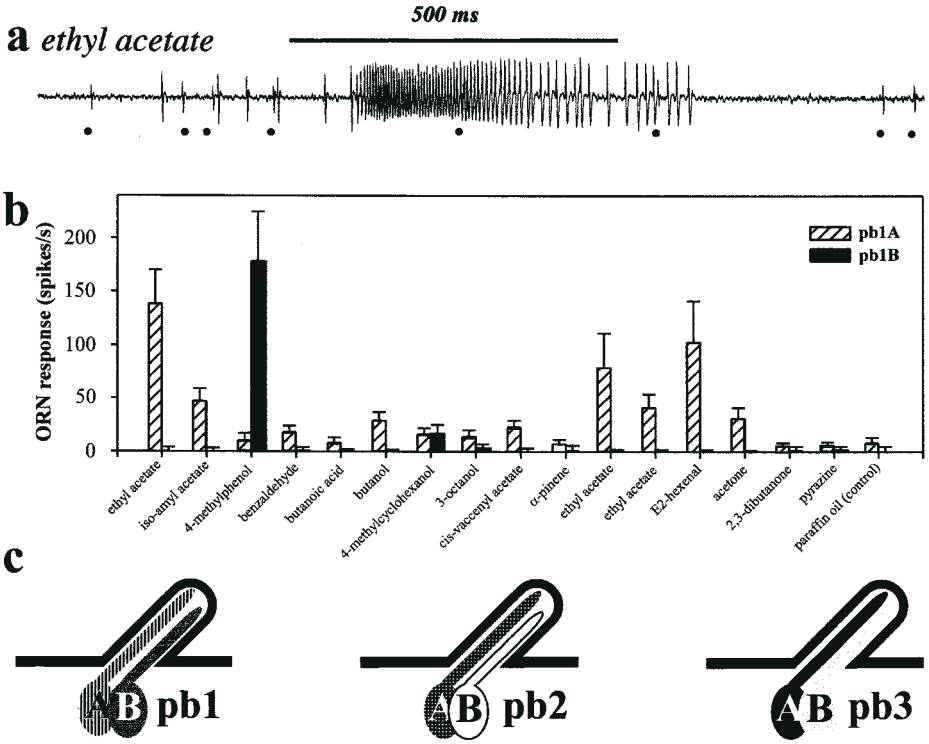

Figure 1 **(a)** A recording from a sensillum (pb1) stimulated with ethyl acetate for 500 ms (horizontal line). Large action potentials, from the A neuron, show an increase in frequency following stimulation with ethyl acetate. Dots indicate smaller action potentials from the B neuron, which is not excited by ethyl acetate. For odor stimulation, air was expelled from a 5 ml syringe over filter paper laden with 20 µl of odorant. The odorant was diluted 10^{-2} in paraffin oil. We do not know the concentration of the odor present in the air reaching the preparation. **(b)** Responses of two neurons (pb1A and pb1B) to a panel of odorants. Error bars indicate SD; $n = 13$. pb1B responds strongly only to one of the tested odors. Responses of pb1A to several odors are significantly greater than the response to the paraffin oil diluent alone. The indicated ORN response is measured as the increase in spikes per second over the spontaneous frequency. **(c)** The six classes of ORN in the maxillary palp, paired in three functional types of sensillum. Adapted from de Bruyne *et al.* (de Bruyne *et al.*, 1999).

(Singh and Nayak, 1985; Shanbhag *et al.*, 2000). The two neurons in a sensillum observe a strict pairing rule, with each sensillum combining neurons of two particular classes. Thus there are three functional types of sensilla, which we term pb1, pb2 and pb3, containing among them six classes of neurons: pb1A, pb1B, pb2A, pb2B, pb3A and pb3B (Figure 1c).

This work has provided some understanding of the cellular basis of odor coding in the maxillary palp. In order to address the molecular basis of coding, we sought to identify the receptor molecules that underlie these cellular characteristics.

Identification of odor receptor genes by e-genetics

As a means of identifying odor receptors and other chemosensory receptors in *Drosophila*, we have used a bioinformatics approach that might be termed 'e-genetics'. Specifically, we made the assumption that odor receptors in *Drosophila* were G-protein-coupled receptors (GPCRs), an assumption based in part on data concerning olfactory transduction in insects (Hildebrand and Shepherd, 1997), and in part on analogy to odor receptors in other organisms (Buck and Axel, 1991; Troemel *et al.*, 1995). Since GPCRs can have widely divergent sequences, but most share a characteristic structure containing seven transmembrane domains, we devised a strategy to scan the *Drosophila* genome database for proteins with particular structures, as opposed to proteins with particular sequences.

To search for proteins with structures like those of GPCRs, we developed a computer algorithm that identifies open reading frames (ORFs) from a DNA database and then maps the predicted proteins into an *n*-dimensional protein space (Clyne *et al.*, 1999b; Kim *et al.*, 2001). The construction of this protein space is critical: our goal was to design a space in which GPCRs, indicated by the solid dots in Figure 2, would be segregated in one region of the space. The key idea was to construct a space that allows interpolation, i.e. if a newly identified protein maps within a region of the space occupied by previously identified GPCRs, then the new protein will probably be a GPCR.

To develop the algorithm we used a training set of previously identified GPCRs and non-GPCRs, eventually using 750 GPCRs and 1000 non-GPCRs extracted from the

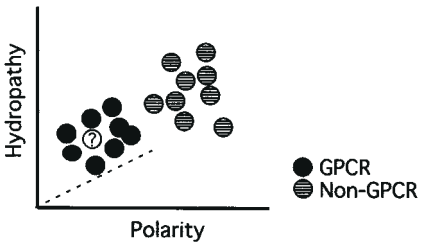

Figure 2 Constructing an *n*-dimensional protein space that allows interpolation. For ease of illustration, a three-dimensional space is shown. Tested variables, represented by dimensions in the space, included hydropathy, polarity, p*I*, pK_a, molecular weight and amino acid composition.

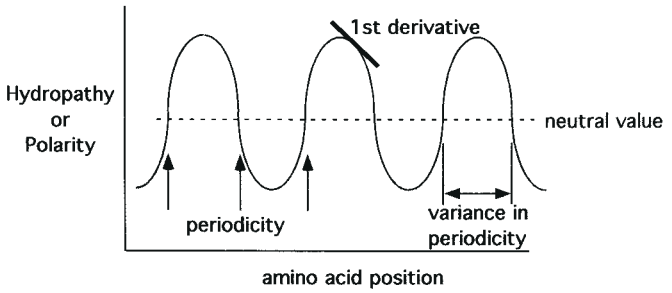

Figure 3 Refining the parameters with a sliding window recognizer. Selected parameters were: average periodicity of the hydropathy function, a measure of the frequency with which the hydropathy function crossed a neutral value; average periodicity of the polarity function; variance in the periodicity of the polarity function; variance in the first derivative of the polarity function; amino acid usage index. A further description of these parameters is given elsewhere (Kim *et al.*, 2001).

SwissProt protein database. To construct a useful space in which GPCRs and non-GPCRs would be distinguishable, we tested a number of different parameters. For ease of illustration, Figure 2 shows a three-dimensional space, with two dimensions labeled as hydropathy and polarity, but in developing the algorithm we used an *n*-dimensional space, and we also tested p*I*, pK_a, molecular weight and amino acid composition, among other parameters.

More specifically, we tested numerous variations of these parameters, refining them in a variety of ways in an attempt to describe protein structure informatively. In selecting variations to test, we based our approach on the fact that GPCRs have multiple transmembrane domains; thus, the physical properties of the polypeptides alternate in a periodic way. For example, with respect to hydropathy there are alternating stretches of hydrophobic and hydrophilic residues. We used a sliding window recognizer (Engelman *et al.*, 1986; von Heijne, 1992) to describe this alternation, and Figure 3 shows that the local hydropathy alternates as a function of amino acid position, ranging from regions of high hydropathy to low hydropathy within a portion of an idealized GPCR protein.

Among 70 different variables tested, we identified five that together were especially useful in distinguishing GPCRs from other proteins (Figure 3). These variables were: (i) the average periodicity of the hydropathy function, which is a measure of the frequency with which the function crosses a neutral value; (ii) the average periodicity of a polarity function, which is related to the hydropathy function; (iii) the variance in the periodicity of the polarity function; (iv) the variance in the first derivative of the polarity function; and (v) an amino acid usage index. We note that the algorithm has evolved since its inception, and some versions use subsets of these variables.

Having thus defined a useful protein space, we next needed to identify a function that provided maximal separation of GPCRs from non-GPCRs within the space. We used a non-parametric linear discriminant function (Gnanadesikan, 1977) to achieve this goal. In a three-dimensional protein space such as that shown in Figure 2, this function can be thought of as a plane that best separates the two sets of proteins. The function is then used to predict whether unknown proteins are GPCRs or non-GPCRs.

How well does this algorithm distinguish GPCRs from non-GPCRs? First we tested another, independent set of 100 known GPCRs and 100 non-GPCRs, which we had previously extracted from the SwissProt database. When we mapped these proteins into the protein space and classified them with the discriminant function, we found that the algorithm correctly predicted 96% of the GPCRs to be GPCRs. None of the non-GPCRs were predicted to be GPCRs; that is, the false positive rate was 0. However, these were full-length protein sequences. In the *Drosophila* genome, most coding regions are broken into introns and exons. Thus, when searching through genomic sequence, the typical ORF extracted by the algorithm encodes only a portion of a protein. (Similarly, in EST databases only a small portion of a protein sequence is available.) Therefore, as a stringent test of the algorithm's ability to classify short ORFs, we tested 100 amino acid stretches of both GPCRs and non-GPCRs and found that, even with such limited amounts of sequence, the algorithm correctly identified >90% of GPCRs and only misidentified 4% of the non-GPCRs as GPCRs.

Having thus developed an algorithm capable of identifying GPCRs from DNA databases, we used it to identify GPCRs in an early release of sequence from the *Drosophila* genome database. The algorithm identified a list of candidate ORFs, in ~1 min of computer time. Some of these ORFs corresponded to known multitransmembrane domain proteins, including some previously identified receptors, channels and transporters. Most of the ORFs contained only a small number of transmembrane domains; we were seeking genes encoding seven transmembrane domains. As suggested above, since we were analyzing a genomic DNA sequence, it seemed likely that some of these ORFs might be spliced to neighboring ORFs, with the spliced

product encoding seven transmembrane domains. Accordingly, for some ORFs we were able to use the consensus fly intron/exon splice sequences (Mount *et al.*, 1992) to identify nearby ORFs, which when spliced together would generate products with approximately seven transmembrane domains.

To determine whether any of the genes encoding novel seven-transmembrane-domain proteins were expressed in the olfactory organs, we designed polymerase chain reaction (PCR) primers that spanned introns in several of them, and using reverse transcriptase (RT)-PCR found that one of the first ones on the list was expressed in the antenna, but no other tested tissue in the fly. We subsequently found that this gene defined a large family of genes, the Or genes, which are expressed in the olfactory system and which are predicted to encode seven-transmembrane-domain proteins (Clyne *et al.*, 1999b). The genes are dispersed widely throughout the genome. At the same time, many are found in clusters of up to three genes. Individual genes of this family are expressed in subsets of ORNs (Figure 4a), and different members are expressed in different subsets (Clyne *et al.*, 1999b; Vosshall *et al.*, 1999).

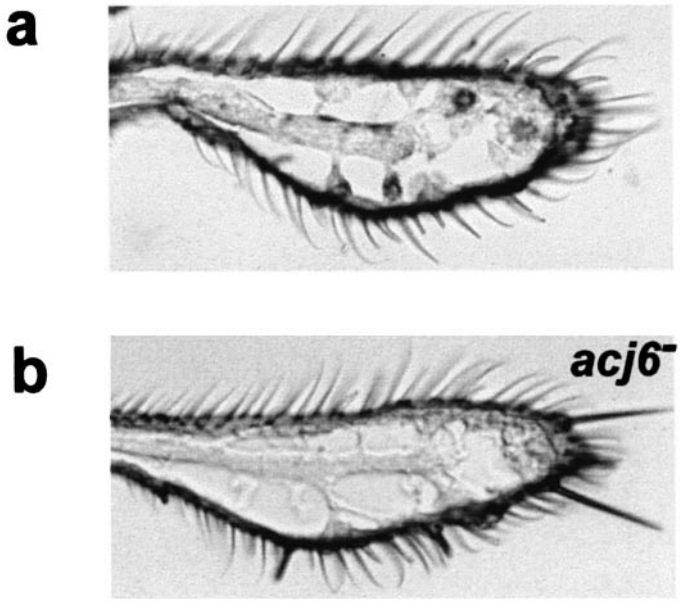

Figure 4 An Or gene is expressed in a subset of ORNs in the wild type **(a)**, but is not expressed in the *acj6*^{6} null mutant **(b)**. *In situ* hybridizations are to sagittal tissue sections of the maxillary palp, the simpler of the fly's two olfactory organs. Note in (a) that unlabeled ORNs are visible under the cuticular surface (top center). Adapted from Clyne *et al.* (Clyne *et al.*, 1999b). © 1999 by Cell Press.

acj6 and the problem of receptor gene choice

The e-genetics we used has thus identified a large family of candidate odor receptor genes in *Drosophila*. Computational identification of these genes was efficient and economical. However, our ultimate goal in receptor identification has been to address central problems in olfactory system function and development, and for this purpose it is useful to take advantage of the power of *Drosophila* genetics.

An intriguing problem in olfaction is how individual neurons choose, from among a large repertoire, which receptors to express. Moreover, how are the choices of individual neurons coordinated so as to produce a system capable of supporting coherent coding? Remarkably little is known about these problems in the olfactory system of any organism.

We isolated an olfactory mutant, *acj6* (*abnormal chemosensory jump 6*), in a screen for mutants defective in an olfactory-driven behavioral response (McKenna *et al.*, 1989). Electroantennogram and electropalpogram recordings showed that the physiological response of the olfactory organs is severely reduced to some, but not all, odors (Ayer and Carlson, 1991). Single-unit physiology subsequently showed that in null mutants of *acj6*, some receptor neurons are normal, some have lost response to all odors and some have undergone alterations in odor specificity (Clyne *et al.*, 1999a). Particularly interesting is that some neurons acquire a novel odor specificity that is different from any we have observed in the wild type. Genetic mapping showed that the *acj6* gene is localized to region 13C1–3 on the X chromosome. Molecular analysis revealed that *acj6* encodes a POU domain transcription factor (Clyne *et al.*, 1999a). Other POU genes have previously been shown to act in specifying the identities of subsets of visual, auditory, somatosensory and mechanosensory neurons in other species (McEvilly and Rosenfeld, 2000).

How might Acj6 act in determining the odor specificity of a subset of ORNs? The simplest model for how it might do so is by regulating a subset of odor receptor genes. We have found this to be the case: a subset of Or genes (8/16 tested) are not expressed in an *acj6* null mutant (Figure 4b) (Clyne *et al.*, 1999b) (also P.J. Clyne *et al.*, unpublished results). Thus we have found that the Acj6 POU domain transcription factor plays a role in the process by which ORNs choose which receptor genes to express. Moreover, these results support the identity of the Or genes as odor receptor genes: one would expect that, in a mutant in which some odor receptors are not expressed normally, some ORNs would have abnormal odor sensitivities, which is observed in mutants of *acj6*.

Why is *acj6* required for the regulation of only a subset of odor receptor genes, and for the determination of odor specificity in only a subset of ORNs? One possible explanation is suggested by the observation that *acj6* is one of several POU genes in *Drosophila*; four POU genes have been reported, all of which are expressed in neurons (Billin *et al.*, 1991; Dick *et al.*, 1991; Lloyd and Sakonju, 1991; Treacy *et al.*, 1991; Yang *et al.*, 1993; Anderson *et al.*, 1995; Poole, 1995). In other organisms, some POU transcription factors have been shown to form homodimers or heterodimers

(Voss *et al.*, 1991). We have found evidence that at least some of the other *Drosophila* POU genes are expressed in the olfactory system, as determined by *in situ* hybridization, immunohistochemistry and/or RT-PCR amplification (unpublished results). These results support the possibility that POU genes act in concert, perhaps combinatorially, to help govern the odor specificity of many or all ORNs in the fly.

Acknowledgements

Supported by grants from the NIH (DC02174) and the Human Frontiers Science Program, and by a McKnight Investigator Award to J.C.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Cocayne, J.D., Amanatides, P.G. *et al.* (2000) *The genomic sequence of* Drosophila melanogaster. Science, 287, 2204–2216.

Anderson, M., Perkins, G., Chittick, P., Shrigley, R. and **Johnson, W.** (1995) drifter*, a* Drosophila *POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia*. Genes Dev., 9, 123–137.

Ayer, R.K. and **Carlson, J.** (1991) acj6*: a gene affecting olfactory physiology and behavior in* Drosophila. Proc. Natl Acad. Sci. USA, 88, 5467–5471.

Billin, A., Cockerill, K. and **Poole, S.** (1991) *Isolation of a family of* Drosophila *POU domain genes expressed in early development*. Mech. Dev., 34, 75–84.

Boeckh, J. and **Ernst, K.-D.** (1987) *Contribution of single unit analysis in insects to an understanding of olfactory function.* J. Comp. Physiol. A: Sens. Neural Behav. Physiol., 161, 549–565.

Buck, L. and **Axel, R.** (1991) *A novel multigene family may encode odorant receptors: a molecular basis for odor recognition*. Cell, 65, 175–187.

Carlson, J. (1996) *Olfaction in* Drosophila*: from odor to behavior*. Trends Genet., 12, 175–180.

Clyne, P., Certel, S., de Bruyne, M., Zaslavsky, L., Johnson, W. and **Carlson, J.** (1999a) *The odor specificities of a subset of olfactory receptor neurons are governed by Acj6, a POU domain transcription factor*. Neuron, 22, 339–347.

Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J.H. and **Carlson, J.R.** (1999b) *A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in* Drosophila. Neuron, 22, 327–338.

de Bruyne, M., Clyne, P.J. and **Carlson, J.R.** (1999) *Odor coding in a model olfactory organ: the* Drosophila *maxillary palp*. J. Neurosci., 19, 4520–4532.

Dick, T., Yang, X., Yeo, S. and **Chia, W.** (1991) *Two closely linked* Drosophila *POU domain genes are expressed in neuroblasts and sensory elements*. Proc. Natl Acad. Sci. USA, 88, 7645–7649.

Engelman, D.M., Steitz, T.A. and **Goldman, A.** (1986) *Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins*. Annu. Rev. Biophys. Biophys. Chem., 15, 321–353.

Gnanadesikan, R. (1977). Methods for Statistical Data Analysis of Multivariate Observations. John Wiley & Sons, New York.

Hildebrand, J.G. and **Shepherd, G.M.** (1997) *Mechanisms of olfactory discrimination: converging evidence for common principles across phyla*. Annu. Rev. Neurosci., 20, 595–631.

Kaissling, K.-E. (1995) *Single unit and electroantennogram recordings in insect olfactory organs*. In Spielman, A.I. and Brand, J.G. (eds), Experimental Cell Biology of Taste and Olfaction, Current Techniques and Protocols. CRC Press, Boca Raton, FL, pp. 361–377.

Kim, J., Moriyama, E., Warr, C., Clyne, P. and **Carlson, J.** (2001) *Identification of novel multi-transmembrane proteins from genomic databases using quasi-periodic structural properties*. Bioinformatics, 16, 767–775.

Lloyd, A. and **Sakonju, S.** (1991) *Characterization of two* Drosophila *POU domain genes, related to oct-1 and oct-2, and the regulationof their expression patterns*. Mech. Dev., 36, 87–102.

McEvilly, R.J. and **Rosenfeld, M.G.** (2000) *The role of POU domain proteins in the regulation of mammalian pituitary and nervous system development*. Prog. Nucleic Acid Res. Mol. Biol., 63, 223–255.

McKenna, M., Monte, P., Helfand, S.L., Woodard, C. and **Carlson, J.** (1989) *A simple chemosensory response in* Drosophila *and the isolation of acj mutants in which it is affected*. Proc Natl Acad Sci USA, 86, 8118–8122.

Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O. and **Fields, C.** (1992) *Splicing signals in* Drosophila*: intron size, information content, and consensus sequences*. Nucleic Acids Res., 20, 4255–4262.

Poole, S. (1995) *Conservation of complex expression domains of the pdm-2 POU domain gene between* Drosophila virilis *and* Drosophila melanogaster. Mech. Dev., 49, 107–116.

Rubin, G. M., Yandell, M.D., Wortman, J.R., Miklos, G.L.G., Nelson, C.R., Hariharan, I. K. *et al.* (2000) *Comparative genomics of the eukaryotes*. Science, 287, 2185–2196.

Schneider, D. and **Boeckh, J.** (1962) *Rezerpotential und nervenimpulse einzelner olfaktorischer sensillen der insektenantenne*. Z. Vergl. Physiol., 45, 405–412.

Shanbhag, S.R., Muller, B. and **Steinbrecht, R.A.** (2000) *Atlas of olfactory organs of* Drosophila melanogaster*. 1. Types, external organization, innervation and distribution of olfactory sensilla*. Int. J. Insect Morphol. Embryol., 28, 377–397.

Siddiqi, O. (1991) *Olfaction in* Drosophila. In Wysocki, C. and Kare, M. (eds), Chemical Senses. Marcel Dekker, New York, pp. 79–96.

Singh, R.N. and **Nayak, S.V.** (1985) *Fine structure and primary sensory projections of sensilla on the maxillary palp of* Drosophila melanogaster *meigen (Diptera: Drosophilidae)*. Int. J.Insect Morphol. Embryol., 14, 291–306.

Stocker, R. (1994) *The organization of the chemosensory system in* Drosophila melanogaster*: a review*. Cell Tissue Res., 275, 3–26.

Treacy, M., He, X. and **Rosenfeld, M.** (1991) *I-POU: a POU-domain protein that inhibits neuron-specific gene activation*. Nature, 350, 577–584.

Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H. A. and **Bargmann, C.I.** (1995) *Divergent seven transmembrane receptors are candidate chemosensory receptors in* C. elegans. Cell, 83, 207–218.

von Heijne, G. (1992) *Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule*. J. Mol. Biol., 225, 487–494.

Voss, J.W., Wilson, L. and **Rosenfeld, M.G.** (1991) *POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can*

cooperate to induce expression of the prolactin promoter. Genes Dev., 5, 1309–1320.

Vosshall, L.B., Amrein, H., Morozov, P.S., Rzhetsky, A. and **Axel, R.** (1999) *A spatial map of olfactory receptor expression in the* Drosophila *antenna*. Cell, 96, 725–736.

Yang, X., Yeo, S., Dick, T. and **Chia, W.** (1993) *The role of a* Drosophila *POU homeo domain gene in the specification of neural precursor cell identity in the developing embryonic central nervous system*. Genes Dev., 7, 504–516.

Accepted October 31, 2000