

Odorant Receptor C-Terminal Motifs in Divergent Insect Species

Authors: Miller, Raymond, and Tu, Zhijian

Source: Journal of Insect Science, 8(53) : 1-10

Published By: Entomological Society of America

URL: <https://doi.org/10.1673/031.008.5301>

The BioOne Digital Library (<https://bioone.org/>) provides worldwide distribution for more than 580 journals and eBooks from BioOne's community of over 150 nonprofit societies, research institutions, and university presses in the biological, ecological, and environmental sciences. The BioOne Digital Library encompasses the flagship aggregation BioOne Complete (<https://bioone.org/subscribe>), the BioOne Complete Archive (<https://bioone.org/archive>), and the BioOne eBooks program offerings ESA eBook Collection (<https://bioone.org/esa-ebooks>) and CSIRO Publishing BioSelect Collection (<https://bioone.org/csiro-ebooks>).

Your use of this PDF, the BioOne Digital Library, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Digital Library content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne is an innovative nonprofit that sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



Odorant receptor c-terminal motifs in divergent insect species

Raymond Miller^{1,a} and Zhijian Tu^{1,b}

¹Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Abstract

Insect odorant receptors are a large family of seven transmembrane proteins believed to be G-protein coupled receptors. The peptide sequences of two odorant receptors within a given species may share as little as 17% identity, and there is limited similarity between receptors of divergent species. One exception is DmOr83b, which is found in *Drosophila melanogaster* and is highly conserved in at least ten other insect species. DmOr83b is broadly expressed in most of the olfactory sensory neurons of *D. melanogaster* at most developmental stages, while other odorant receptors tend to have more restricted and specific expression patterns. DmOr83b is critical for *D. melanogaster* olfaction, and it is involved in properly localizing other odorant receptors possibly by forming heterodimers with these receptors. The C-terminal region has been implicated as sites for such heterodimer formation. Multiple em for motif elicitation (MEME), a hidden markov model based program, was used to uncover three conserved motifs in the C-termini of a vast majority of the odorant receptor peptides from *Anopheles gambiae*, *D. melanogaster*, and *Apis mellifera*. These motifs are also found in DmOr83b and its orthologs and the order of these motifs is conserved as well. The conservation of these motifs among divergent odorant receptors in divergent species suggests functional importance. We propose that these motifs are involved in receptor-receptor protein interactions, contributing to the heterodimer formation between DmOr83b (or its orthologs) and other odorant receptors.

Keywords: mosquito, G-protein coupled receptor, MEME, heterodimer, hidden markov model, Or83b

Abbreviations: MEME - multiple em for motif elicitation, MAST - motif alignment and search tool, DmOr - *Drosophila melanogaster* odorant receptor, AgOr - *Anopheles gambiae* odorant receptor, AmOr - *Apis mellifera* odorant receptor

Correspondence: ^aramiller@vt.edu ^bjaketu@vt.edu

Received: 26 August 2007 | **Accepted:** 1 October 2007 | **Published:** 22 September 2008

Copyright: This is an open access paper. We use the Creative Commons Attribution 3.0 license that permits unrestricted use, provided that the paper is properly attributed.

ISSN: 1536-2442 | Volume 8, Number 53

Cite this paper as:

Miller R, Tu Z. 2008. Odorant receptor c-terminal motifs in divergent insect species. *Journal of Insect Science* 8:53, available online: insectscience.org/8.53

Introduction

Insect olfaction and olfactory signaling is a rapidly growing area of research (Rutzler and Zwiebel 2005). Several protein families are being studied that include odorant binding proteins, sensory neuron membrane proteins, odorant degrading enzymes, and odorant receptors. A large body of recent literature has been written on insect odorant receptors (Clyne et al. 1999; Vosshall et al. 1999; Hill et al. 2002; Sakurai et al. 2004; Robertson and Wanner 2006). Most of the papers addressing insect odorant receptors report either the discovery of receptor genes in an insect species (Hill et al. 2002; Robertson et al. 2003), and/or the expression of selected odorant receptor genes at various points of the organism's life cycle (Melo et al. 2004). Odorant receptor gene expression is usually localized to the insect sensory organs such as antenna and maxillary palp (Vosshall et al. 2000; Fox et al. 2001), and more recently have been found to be expressed in the proboscis (Kwon et al. 2006). There are also a growing number of papers addressing the specific functions of several odorant receptor proteins (Wetzel et al. 2001; Hallem et al. 2004; Larsson et al. 2004; Sakurai et al. 2004).

Insect odorant receptors have been reported to be putative G-protein coupled receptors (Clyne et al. 1999; Gao and Chess 1999; Hill et al. 2002), but recently this status has been questioned (Benton 2006; Wistrand et al. 2006). The most extensively researched insect odorant receptor is DmOr83b in *Drosophila melanogaster*. A highly conserved ortholog of DmOr83b has been found in all insect species with sufficient genomic sequence information. This list includes *D. melanogaster* (Vosshall et al. 2000), *Anopheles gambiae* (Pitts et al. 2004), *An. stephensi* (R. Miller and Z. Tu, unpublished data), *An. quadrimaculatus* (R. Miller and Z. Tu, unpublished data), *Aedes aegypti* (Melo et al. 2004), *Culex quinquefasciatus* (Xia and Zwiebel 2006), *Bombyx mori* (Sakurai et al. 2004), *Heliothis virescens* (Krieger et al. 2002), *Apis mellifera* (Robertson and Wanner 2006), and *Tribolium castaneum* (GenBank Accession XP_973196. Note that the GenBank name for OR is GPROR. This is in contrast to the vast majority of the other insect odorant receptors, which are not conserved between species of different genera. DmOr83b is broadly expressed in most of the olfactory sensory neurons of *D. melanogaster* at most stages of development (Vosshall et al. 2000; Larsson et al. 2004). This again is in contrast to other odorant receptors, which have been reported to have a restrictive expression pattern (Vosshall et al. 2000; Fox et al. 2001). *D. melanogaster* lacking a copy of DmOr83b are not able to respond to olfactory cues, and other odorant receptors are not properly localized to the membrane of olfactory sensory neurons (Larsson et al. 2004). DmOr83b is capable of forming a heterodimer with at least one *D. melanogaster* odorant receptor: DmOr43a (Neuhaus et al. 2005). The requirement of a

heterodimer of two G-protein coupled receptors has only been previously observed in the GABA complex where heterodimer formation is required for the function of potassium/calcium channels (Jones et al. 1998; White et al. 1998). Benton and co-authors provides further evidence of heterodimer formation involving DmOr83b with DmOr22a/b, and additionally point to the C-terminal domain of odorant receptor peptides as being the site of heterodimer formation (Benton et al. 2006). The specific location(s) of the protein-protein interaction(s) were not explored. However, previous reports have indicated limited amino acid conservation occurring in the C-terminal end of *D. melanogaster* odorant receptor (DmOr) peptides, including a nearly invariable tryptophan residue (Clyne et al. 1999; Scott et al. 2001; Vosshall 2003).

Using a hidden markov model based program called multiple em for motif elicitation (MEME) (Bailey and Elkan 1994), we have discovered three C-terminal motifs in 76 of the 79 previously annotated *An. gambiae* odorant receptor peptides (Hill et al. 2002). Subsequent analysis indicates that these motifs are conserved within the odorant receptor peptides of *D. melanogaster* and *Ap. mellifera* (Robertson and Wanner 2006). This is significant given that it has been reported that insect odorant receptor peptides are highly divergent within and between species (Clyne et al. 1999; Vosshall et al. 1999; Hill et al. 2002; Vosshall 2003). For example, the amino acid identity between insect odorant receptors of the same species is only 17% in some cases (Vosshall 2003). We hypothesize that these motifs are protein-protein interaction sites involved in odorant receptor-odorant receptor interactions or potentially heterodimer formation between DmOr83b and other odorant receptors.

Materials and Methods

Alignment of *An. gambiae* odorant receptor peptides with ClustalW

All 79 *An. gambiae* odorant receptor peptides (Hill et al. 2002) were aligned using ClustalW v1.83.1 (Thompson et al. 1994). Default parameters were used (multiple alignment gap opening penalty = 10, gap extension penalty = 0.2). Alignments were illustrated using the Jalview java alignment editor (Clamp et al. 2004).

Motif discovery in odorant receptors peptides using MEME

An. gambiae and *Ap. mellifera* odorant receptor peptide sequences were obtained from the supplementary material of two separate studies (Hill et al. 2002; Robertson and Wanner 2006). Fifty-nine *D. melanogaster* odorant receptor peptides were obtained from the Ensembl database (<http://www.ensembl.org>) and were used in the analysis. The program multiple em for motif elicitation (MEME) (Bailey and Elkan 1994) (<http://meme.sdsc.edu/meme/>) version 3.5.1 was compiled on a Macintosh computer

running Mac OS 10.4.8. Each MEME analysis was run with peptide dataset from each species as input. For all three datasets MEME was run using the following command line: *meme dataset_name -protein -mod zoops -minw 15 -maxw 45 -wg 8 -ws 0.2 -evt .00001 -nmotifs 8*. The program command call is *meme*, while *dataset_name* identifies the input dataset, *-protein* indicates the dataset contained peptide sequences, and *-mod* defines the search model. The remaining parameters were *-minw*, which sets the minimum possible motif width at 15 residues, *-maxw*, which sets the maximum possible motif width at 45 residues, *-wg*, which is the gap opening penalty, *-ws*, which is the gap extension penalty, *-evt*, which is the maximum e-value for a motif to be reported, and *-nmotifs*, which indicated the number of motifs that are searched for in the input dataset. Gap opening and extension penalties were reduced from the default values of *wg*=11 and *ws*=1 to *wg*=8 and *ws*=0.2 to reduce artificial breakup of the motifs due to small insertions or deletions. In addition to searching for the top 3 motifs as set by default, *-nmotifs 8* was used to determine whether more than three motifs existed in each dataset.

MAST searching of *An. gambiae* gustatory receptor peptides for odorant receptor motifs

The motif alignment and search tool (MAST) (Bailey and Gribskov 1998), another program in the MEME package, was used to search for AgOr motifs in all 76 *An. gambiae* GRs (Hill et al. 2002). MAST version 3.5.1 was installed as part of the MEME package (see above). Command line used for MAST was: *mast motif_matrices_found_by_meme -d database_of_AgGr*s. The *motif_matrices_found_by_meme* are the profile matrices of the motifs found in a previous MEME analysis and they effectively define the motifs. These matrices were used to search the *database_of_AgGr*s, where AgGr stands for *An. gambiae* gustatory receptors. No other parameters were used.

Weblogo diagrams

All weblogo diagrams were constructed using the weblogo program (Crooks et al. 2004) (<http://weblogo.berkeley.edu/>). MEME output includes BLOCKS of the motifs. If an odorant receptor peptide sequence was found to have a motif, the part of the peptide sequence that contains that motif was used in an alignment, which produced an aligned BLOCK. The aligned BLOCK was used to construct weblogos.

Results

ClustalW alignments of *An. gambiae* odorant receptors

An alignment of all 79 *An. gambiae* odorant receptor (AgOr) peptides using the multiple sequence alignment program ClustalW (Thompson et al. 1994) revealed very little strict sequence conservation (Figure 1). There were

a small number of conserved or highly prevalent residues located in the C-terminal region (Figure 1, blue-colored residues). One of these highly conserved residues is a tryptophan residue found in all but four AgOrs. The lack of strict sequence conservation in AgOrs, and the prevalence of the conserved tryptophan residue is consistent with what has been previously reported for DmOr peptides (Vosshall 2003).

MEME identifies c-terminal motifs in *An. gambiae* odorant receptors

To locate conserved patterns a hidden markov model based program named multiple em for motif elicitation (MEME) was used (Bailey and Elkan 1994). MEME has been used to locate potential regulatory sites in sequences upstream of genes (Ohler et al. 2002), potential protein-protein interaction domains (Fang et al. 2005), and homologous genes missed by homology search (Janssen et al. 2004). One key advantage of MEME over common alignment programs is its ability to find motifs that are not absolutely conserved in consensus sequence. Other advantages of MEME are its speed, no need for prior knowledge about a dataset, and its ability to locate motifs that may not be in the same order through all members of a dataset.

All 79 AgOr peptides were used as input for MEME run using a gap opening penalty of eight and a gap extension parameter of 0.2. Three motifs were identified within the dataset all with highly significant e-values ($4.2e^{-401}$, $4.5e^{-367}$, and $1.1e^{-332}$) (Figure 2, Table 1). All three motifs were present within the last 70 or 90 amino acid residues of the C-terminal end of AgOr peptides, and 76 out of 79 (96%) AgOr peptides had all three motifs. The order of the motifs from N-terminal to C-terminal is motif 3, motif 2, and motif 1. MEME numbers the motifs according to their relative e-values with motif 1 having the best e-value. In subsequent discussions, the three motifs are referred to as motif A, motif B, and motif C, with motif A being furthest of the three from the C-terminus and motif C being the closest to the C-terminus. This naming system is used to allow meaningful comparison between results from different species where these motifs have different ranks of e-values relative to each other. The combined p-value of finding all of the identified motifs in a given odorant receptor peptide in the dataset ranged from $2.49e^{-12}$ to $2.62e^{-38}$. The combined p-value was the probability of finding a match of a sequence in the dataset to a group of motifs by random chance (Figure 2). Significantly, AgOr7, the mosquito ortholog of DmOr83b, has all three motifs at the C-terminal end (Figure 2, asterisk). When the number of motifs for MEME was increased to search for from three to eight motifs, only one additional motif was found with a significant distribution ($2.9e^{-295}$, present in 63 of 79 AgOr peptides). This motif had limited sequence conservation with the notable exception of a histidine residue located

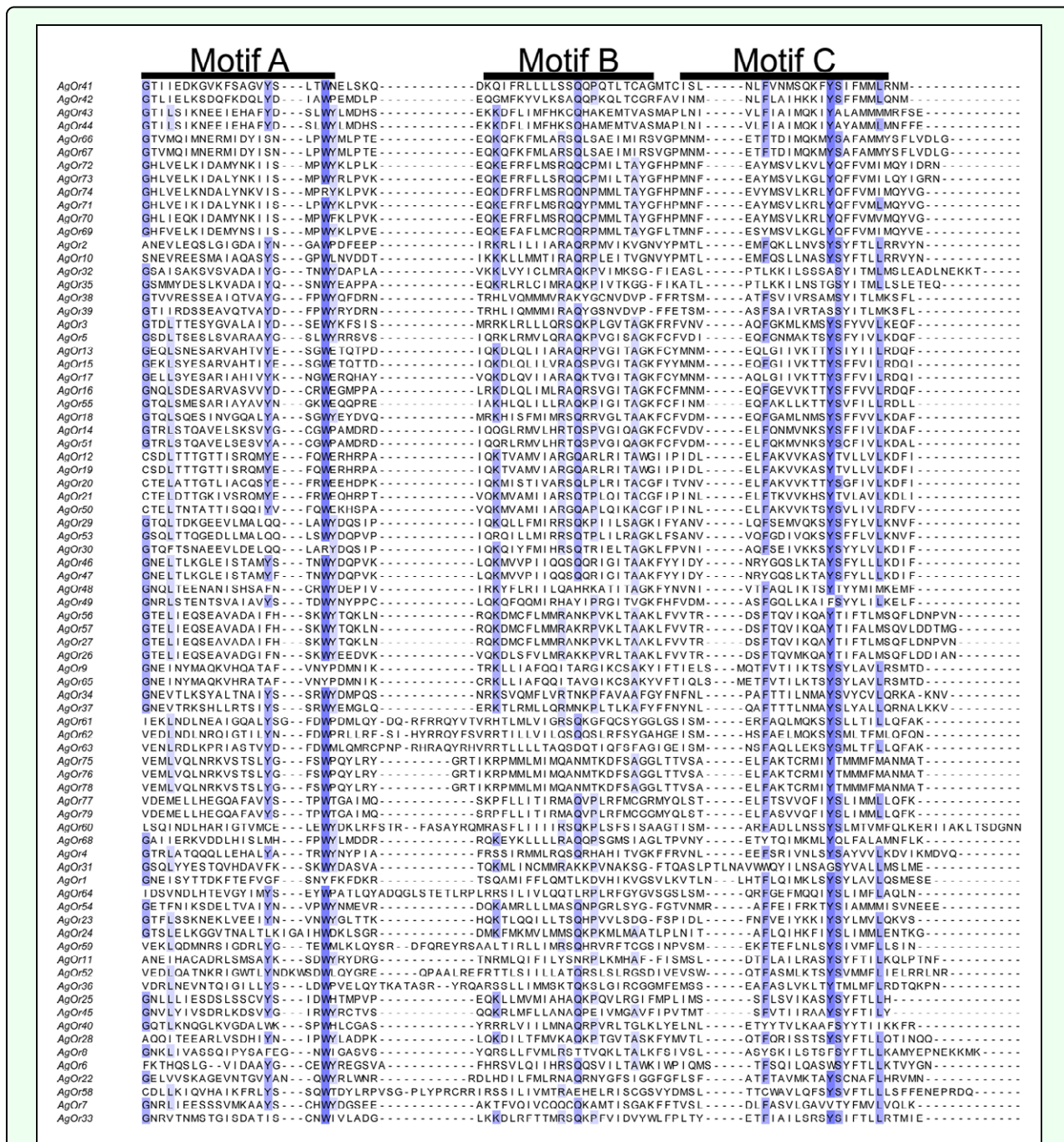


Figure 1. Multiple sequence alignment of the C-terminal region of all 79 *Anopheles gambiae* odorant receptors. AgOr peptides were aligned using ClustalW, and the subsequent alignment visualized using Jalview. A residue present at a given site in 50% or more of the AgOr peptides is boxed in blue. The more intense the blue the more often the residue is found at that site. Only the C-terminal region of the alignment is shown. The positions of motifs A, B, and C are shown. These motifs were not identified using alignment shown here. Instead they were identified using MEME. See Table 1 and Figure 3 for details.

approximately 70 residues to the N-terminal of *An. gambiae* motif A.

Weblogo diagrams of motif A (Figure 3A), motif B (Figure 3B) and motif C (Figure 3C) illustrate the level of amino acid conservation within each motif at each position (Crooks et al. 2004). It is apparent from the weblogo diagrams that only a small portion of each motif consists

of highly prevalent residues although there are additional areas where the chemical properties of the residues such as hydrophobicity, charge, and side chain structure are conserved. For example, in motif B of *An. gambiae* (Figure 3B) residue 1 and 2 are predominately positively charged residues while residues 4, 6, 7, and 8 are hydrophobic. The most highly conserved residues in the AgOr motifs are the tryptophan residue in motif A (Figure 3A), and a

Table 1. Three conserved C-terminal motifs in *An. gambiae*, *D. melanogaster*, and *Ap. mellifera* odorant receptor peptides.

	Motif ¹	No. of odorant receptors containing motif ²	Motif E-value ³	Prevalent Amino Acid Sequence ⁴
<i>An. gambiae</i>	Motif A	78/79	1.1e ⁻³³²	G[T]N]ELTX[K][E][S]N]EX[V]I][A]S]DA[I]L[V]YSSPWY
	Motif B	77/79	4.5e ⁻³⁶⁷	[Q R]KXL[L]R][F]L][M]I][L][I]M]MR[A]S]QKP[L]V]G[I]L]TAG[K]G]
	Motif C	79/79	4.2e ⁻⁴⁰¹	[V]M][N]S][M]L]E[T]L]FAX[I]V][L]V]K[T]K]SYS[Y]F]T[M][V]L][L]M]
<i>D. melanogaster</i>	Motif A	59/59	6.0e ⁻³⁶⁸	[Q]E][L]T]F[P]L]YCY[Y]G[A][N]T]L[L]I]V]XXESEX[L]V]AXA[A]L][Y]F]SSNWY
	Motif B	54/59	1.1e ⁻¹⁷⁷	Y[R]K][R]K]XLL[L]F]F[I]L]M]MR[A]S]Q[R]K]Q]PVX[L]I][K]R]T]JAG
	Motif C	59/59	1.7e ⁻²¹⁴	[S]N][L]M]XTFX[A]S][I]V]L[K]R]X[A]S][Y]F]F[T]A]LL[R]K]SM
<i>Ap. mellifera</i>	Motif A	164/170	6.8e ⁻¹⁰⁹¹	G[Q]E]D][I]L]L[E]D][E]Q][S]C]X[N]E][I]V][G]A]NAVYMSNWY
	Motif B	152/170	3.7e ⁻⁸⁸⁰	[K]L]DLIL[I]V][I]M]IRS[S]N]XPCK[L]I]TAG[K]G][L]I]F]
	Motif C	160/170	4.8e ⁻¹⁰⁸⁶	D[L]M]V]SLETF[T]G]S[I]V][L]I][K]S]T[S]A]FSY[L]F][N]T][L]V]LR

¹The naming of motifs A, B, and C is described in Results section. They are from N- to C- terminus.
²The number of odorant receptors containing a given motif is given as a fraction of the total number of odorant receptors in a species.
³An estimate of the likelihood of each motif being found in the dataset by random chance.
⁴The sequences are from MEME output and they reflect the amino acid residues that are most frequent at these positions. For example, having a G in the output sequence does not mean that all odorant receptors of that species will have a G residue at that position. A [] indicates more than one amino acid residue is frequent at this position.

tyrosine/serine dyad in motif C (Figure 3C) mentioned above.

Odorant receptor c-terminal motifs are not found in gustatory receptors

Insect gustatory receptors are another family of putative G-protein coupled receptors. Insect gustatory receptors and odorant receptors are the closest relatives to each other in evolutionary terms (Clyne et al. 2000; Scott et al.

2001). DmGr21a in *D. melanogaster* is able to confer response to carbon dioxide (Suh et al. 2004) in conjunction with DmGr63a (Jones et al. 2007; Kwon et al. 2007). The motif alignment and search tool (MAST) (Bailey and Gribskov 1998) was used to search for the previously identified AgOr motifs in all 76 *An. gambiae* GRs (Hill et al. 2002). The best hit showed an e-value of 0.033 for a motif in the C-terminal region of one gustatory receptor. The poor e-value of the hit as well as further manual

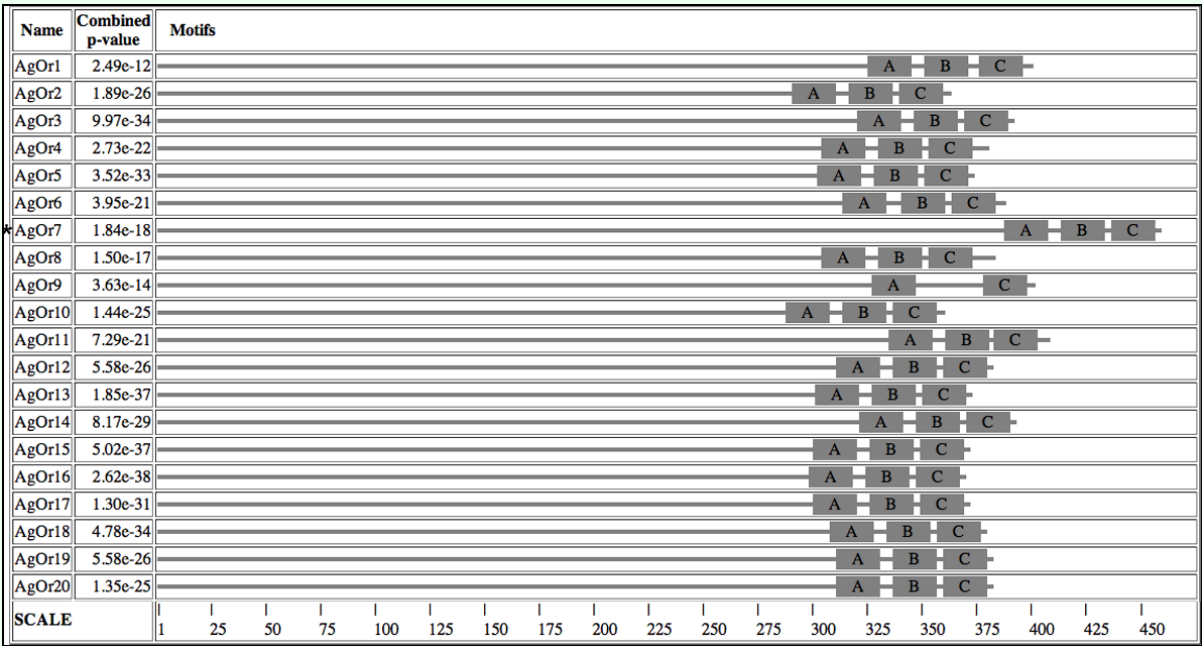


Figure 2. C-terminal motifs found in *Anopheles gambiae* odorant receptors. The image was taken directly from the MEME output and shows the position of three C-terminal motifs located in the first 20 AgOr peptides. Only the first 20 AgOrs were shown to save space. The asterisk points to AgOr7, which is the *An. gambiae* ortholog of DmOr83b. The combined p-value is the probability of finding a match of a sequence in this dataset to a group of motifs by random chance.

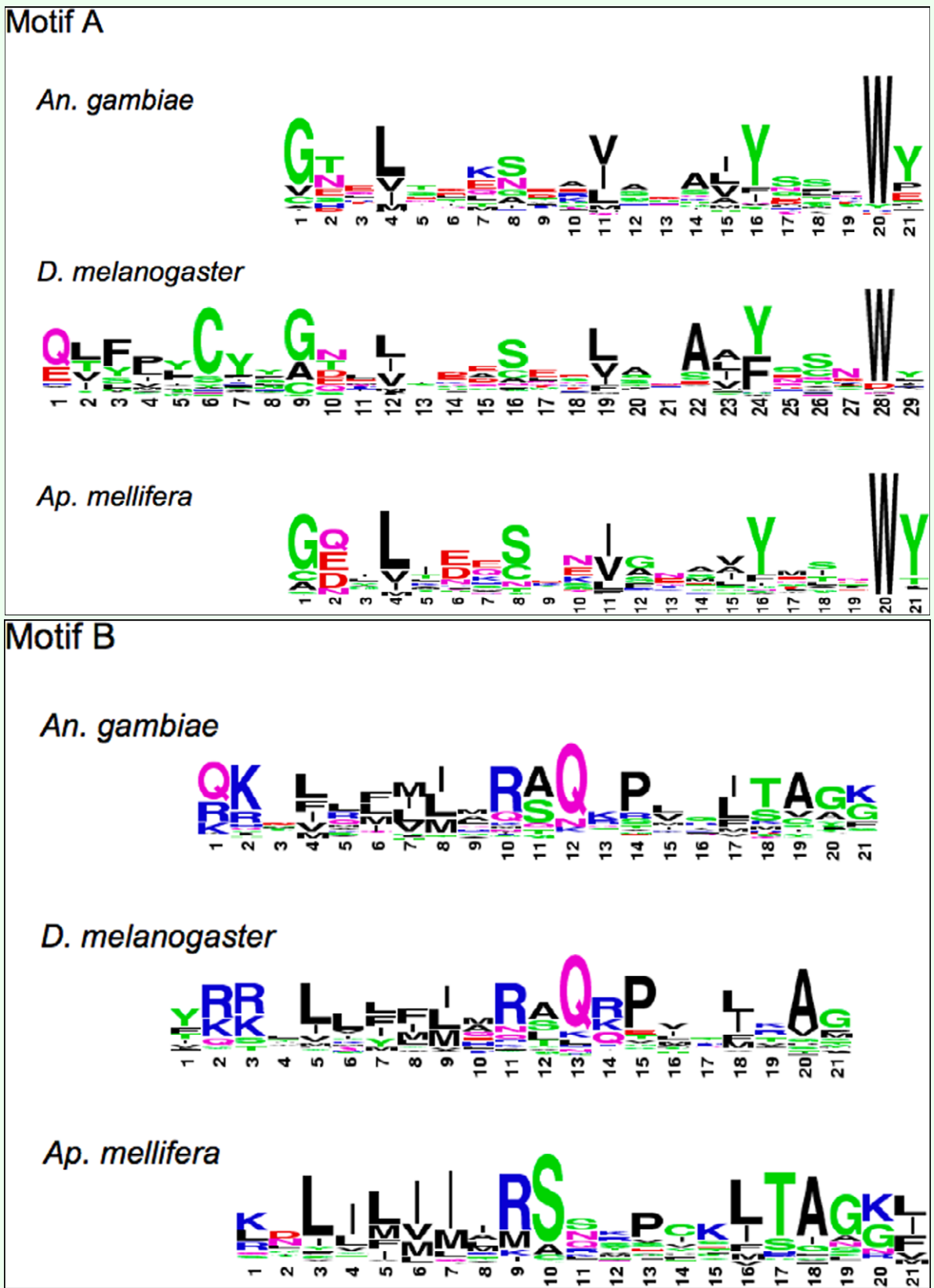
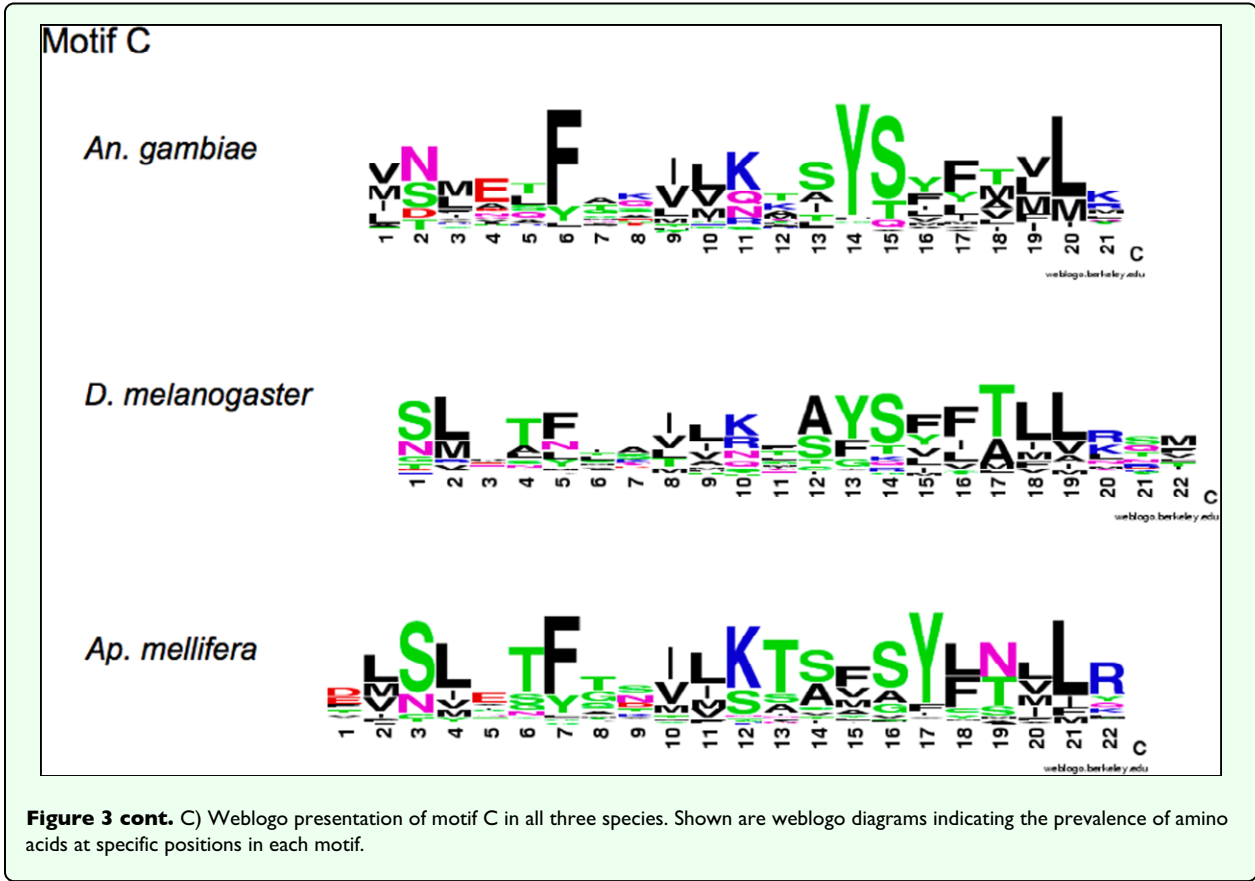


Figure 3. Weblogo presentation of motifs A, B, C in *Anopheles gambiae*, *Drosophila melanogaster*, and *Apis mellifera* odorant receptor peptides. Each line contains weblogo diagrams for motifs A, B, or C in one species. Weblogo diagrams indicate the prevalence of amino acids at specific positions. A). Weblogo presentation of motif A in all three species. B). Weblogo presentation of motif B in all three species.



inspection suggests that it is not a true match. Thus this analysis indicates that the AgOr motifs are specific to odorant receptors and not a feature of G-protein coupled receptors.

Odorant receptor c-terminal motifs are found in *D. melanogaster* and *Ap. mellifera* odorant receptors

A DmOr peptide database of 59 DmOrs was used as input into MEME to determine if any similar motifs existed in these odorant receptors. This analysis revealed three motifs found in the C-terminal end of a vast majority of these peptides (Table 1). All three motifs were found in 54 of 59 (92%) DmOrs. As was the case in AgOr peptides these three motifs are in the same order in all DmOr peptides. A side-by-side comparison of the weblogo diagrams from motifs A and B in both species reveals obvious similarities in sequence (Figure 3A and 3B). Most significant is the highly conserved tryptophan residue in motif A of both species. Part of motif A in DmOr peptides has been previously identified as the sequence of Phe-Pro-X-Cys-Tyr-(X)₂₀-Trp (Vosshall 2003). The analysis showed several additionally conserved residues such as a glycine (residue 9) and a tyrosine/phenylalanine (residue 24). Motif C is very similar in both species in terms of their sequences and boundaries (Figure 3C).

Eight motifs were found in *Ap. mellifera* odorant receptor peptides (AmOr). Three of the motifs are apparent orthologs to the dipteran motifs A, B, and C (Table 1, Figure 3) both in terms of their sequence and relative location. Among the eight AgOr motifs, motifs A, B, and C ranked as number 1, 4, and 2 in terms of the significance of their respective e-values. The motif that had the third best e-value was near the middle of the receptor peptide, and is not shared with the dipteran receptors. Motifs ranked number 5 to 8 appear to have limited distribution in subgroups of AmOr peptides, and thus are not universal motifs in all AmOrs. These motifs are not further discussed in this paper. All three motifs are present in 147 of 170 (86%) AmOr peptides (Table 1, Figure 3). Motifs A, B, and C in AmOrs share similar sequence with dipteran Motifs A, B, and C respectively (Figure 3). For example there is a highly prevalent glycine residue followed by two variable residues, and then a highly prevalent leucine residue in motif A of all three species in addition to the conserved tryptophan residue. AmOr motif C is again very similar to the dipteran motifs (Figure 3C). However, instead of a tyrosine/serine dyad there is a phenylalanine/serine dyad in AmOr. The MEME analysis has therefore found three C-terminal motifs that are located in *An. gambiae*, *D. melanogaster* and *Ap. mellifera* odorant receptor peptides. Most of the residues in these motifs are not highly conserved, but several are highly prevalent across these diverse insect species.

Discussion

Three motifs were located in the C-terminal ends of the odorant receptor peptides of three divergent insect species *An. gambiae*, *D. melanogaster*, and *Ap. mellifera* using a hidden markov model program. Table 1 lists the number of odorant receptors containing these motifs in each species, the e-value of the motifs, and the prevalent amino acid sequences of these motifs. The vast majority of insect odorant receptor peptides analyzed contain these C-terminal motifs. This is interesting considering that insect odorant receptor proteins are a very diverse family having very little conservation between species or within one species (Clyne et al. 1999; Vosshall et al. 1999; Robertson et al. 2003; Vosshall 2003). These motifs were not found in *An. gambiae* GRs despite the close evolutionary relationship between the odorant receptor and GR families (Clyne et al. 2000; Scott et al. 2001).

Although all of the motifs described above had wide distribution in odorant receptors of the three species, motif B was not present in a small, but significant number of odorant receptors, especially in *Ap. mellifera* (Table 1). The absence of motif B may be explained by either technical or biological reasons, or both, as described below. Eleven of the 18 AmOrs lacking motif B had incomplete C-termini in current annotation, and two of the peptides were clearly pseudogenes (Robertson and Wanner 2006). Motif B was also not found in AmOr2, which is the honeybee ortholog of DmOr83b. However, a close inspection of the AmOr2 sequence revealed no amino acid substitution in the motif B region in comparison with DmOr83b and one substitution in comparison with AgOr7 (Figure 4). Therefore, sequence variation between motif B of the three species may explain why nearly identical sequences were recognized as motif B in DmOr83b and AgOr7 but not in AmOr2. Motif B was also lacking in two AgOrs and four DmOrs. Motif B was not as well conserved as the other two motifs (Figure 3). It is possible that the specific sequence of motif B is not as important as the chemical or structural properties of the residues in this motif. In comparing motif B in all three insect species (Figure 3B), some amino acid residues are present that are highly variable, but most of the residues in this region are hydrophobic in character. This conservation of hydrophobicity in these five residues may be functionally significant, while at the same time are difficult to be recognized by computer programs. It is also possible that motif B serves a role in enhancing a

biological process, but is not absolutely required. For example, based on the working hypothesis that these C-terminal motifs are involved in protein-protein interactions, odorant receptor proteins lacking motif B might have a lower binding efficiency.

Having identified these motifs it is appropriate to ask why these motifs are present in the highly diverse insect odorant receptor family? As mentioned above, one possibility is that these motifs are involved in protein-protein interactions. There have been many efforts to identify protein-protein interaction sites through *in silico* methods, which resulted in the identification of several key characteristics. Protein-protein interaction sites are exposed on the surface of proteins and are hydrophobic, circular, and protruding (Janin and Chothia 1990; Young et al. 1994; Jones and Thornton 1995; Jones and Thornton 1997). Within these interaction areas are small “hot-spots” of a few residues contributing greatly to the overall binding energy of protein-protein interactions (Clackson and Wells 1995). In one survey it was found that tryptophan, tyrosine, and arginine are highly prevalent in these “hot-spots” (Bogan and Thorn 1998). Another study reported that tryptophan, phenylalanine, and methionine residues are significantly conserved in binding sites, but not on other exposed surfaces of proteins (Ma et al. 2003). Highly conserved and prevalent tryptophan, tyrosine, phenylalanine, and arginine residues were located in the C-terminal motifs of odorant receptors (Figure 3). We hypothesize that these motifs are protein-protein interaction sites, which would explain the conservation of only a few residues across the highly diverse insect odorant receptor protein family.

Unfortunately, at present there is no X-ray crystal structure of any insect odorant receptor or gustatory receptor that may illuminate the exact positioning of the newly discovered motifs and their potential role in protein-protein interaction. Hydrophobicity analysis can be useful at least in determining where residues are in relation to transmembrane helices. In this study the TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>) was used to predict the transmembrane helices of five odorant receptors each from *An. gambiae*, *D. melanogaster*, and *Ap. mellifera* along with DmOr83b (data not shown). Motif A in all three species was found within helix 6, and perhaps part of the helix 6–7 loop. The difficulty of pinning down the exact positioning of helix 7 in particular makes this conclusion hard to draw unequivocally. It also makes

DmOr83b	QVFHFCIFGNRLIESSSVMEAAYSCHWYDGSEEAFTFVQIVCQCQKAMTISGAKFFTVSLDLFASVLGAVVTFYFMVLVQLK
AgOr7	QVFLFCIFGNRLIESSSVMKAAYSCHWYDGSEEAFTFVQIVCQCQKAMTISGAKFFTVSLDLFASVLGAVVTFYFMVLVQLK
AmOr2	QVFMLCIFGNRLIESSSVMEAAYSCHWYDGSEEAFTFVQIVCQCQKAMTISGAKFFTVSLDLFASVLGAMVTFYFMVLVQLK
	*** : ***** : ***** : ***** : ***** : ***** : *****

Figure 4. Conservation at the C-terminal regions of DmOr83b and its orthologs. Shown here is a ClustalW alignment of the last ~90 amino acid residues of the Or83b family members in *Drosophila melanogaster* (DmOr83b), *Anopheles gambiae* (AgOr7), and *Apis mellifera* (AmOr2). The relative position of motifs A, B, and C are shown.

further analysis of the positions of motif B and C uninformative except the supposition that these motifs lie near or in helix 7.

The vast majority of these insect odorant receptors maintained these motifs across hundreds of millions of years of evolution. This is impressive considering that the identity between insect odorant receptor peptides of the same species in some cases is as low as 17% (Vosshall 2003). These motifs have several highly conserved amino acids that were identified as being important in protein-protein interactions in other models. It is possible that these motifs allow odorant receptor-odorant receptor interactions as has been reported *in vitro* (Neuhaus et al. 2005). A more tantalizing prospect is that all or some of these motifs are involved in the formation of a heterodimer complex between DmOr83b or its ortholog and other odorant receptors (Neuhaus et al. 2005; Benton et al. 2006), a hypothesis that may be tested experimentally.

Acknowledgements

We thank the anonymous reviewers for their comments. This work was supported by a NIH grant AI063252 to Z. Tu.

References

- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of International Conference on Intelligent Systems for Molecular Biology* 2: 28-36.
- Bailey TL, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 14: 48-54.
- Benton R. 2006. On the ORigin of smell: odorant receptors in insects. *Cellular and Molecular Life Sciences* 63: 1579-1585.
- Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS Biology* 4: 240-257.
- Bogan AA, Thorn KS. 1998. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* 280: 1-9.
- Clackson T, Wells JA. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* 267: 383-386.
- Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. *Bioinformatics* 20: 426-427.
- Clyne PJ, Warr CG, Carlson JR. 2000. Candidate taste receptors in *Drosophila*. *Science* 287: 1830-1834.
- Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* 22: 327-338.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* 14: 1188-1190.
- Fang J, Haas RJ, Dong Y, Lushington GH. 2005. Discover protein sequence signatures from protein-protein interaction data. *BMC Bioinformatics* 6: 277.
- Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ. 2001. Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proceedings of the National Academy of Sciences USA* 98: 14693-14697.
- Gao Q, Chess A. 1999. Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* 60: 31-39.
- Hallem EA, Nicole Fox A, Zwiebel LJ, Carlson JR. 2004. Olfaction: mosquito receptor for human-sweat odorant. *Nature* 427: 212-213.
- Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ. 2002. G Protein-Coupled Receptors in *Anopheles gambiae*. *Science* 298: 176-178.
- Janin J, Chothia C. 1990. The structure of protein-protein recognition sites. *Journal of Biological Chemistry* 265: 16027-16030.
- Janssen CS, Phillips RS, Turner CM, Barrett MP. 2004. *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Research* 32: 5712-5720.
- Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, Dai M, Yao WJ, Johnson M, Gunwaldsen C, Huang LY, Tang C, Shen Q, Salon JA, Morse K, Laz T, Smith KE, Nagarathnam D, Noble SA, Branchek TA, Gerald C. 1998. GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* 396: 674-679.
- Jones S, Thornton JM. 1995. Protein-protein interactions: a review of protein dimer structures. *Progress in Biophysics and Molecular Biology* 63: 31-65.
- Jones S, Thornton JM. 1997. Analysis of protein-protein interaction sites using surface patches. *Journal of Molecular Biology* 272: 121-132.
- Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. 2007. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* 445: 86-90.
- Krieger J, Raming K, Dewar YM, Bette S, Conzelmann S, Breer H. 2002. A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. *European Journal of Neuroscience* 16: 619-628.
- Kwon HW, Lu T, Rutzler M, Zwiebel LJ. 2006. Olfactory responses in a gustatory organ of the malaria vector mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences USA* 103: 13526-13531.
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO₂ reception in *Drosophila*. *Proceedings of the National Academy of Sciences USA* 104: 3574-3578.
- Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. 2004. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* 43: 703-714.
- Ma B, Elkayam T, Wolfson H, Nussinov R. 2003. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences USA* 100: 5772-5777.

- Melo AC, Rutzler M, Pitts RJ, Zwiebel LJ. 2004. Identification of a chemosensory receptor from the yellow fever mosquito, *Aedes aegypti*, that is highly conserved and expressed in olfactory and gustatory organs. *Chemical Senses* 29: 403-410.
- Neuhaus EM, Gisselmann G, Zhang W, Dooley R, Stortkuhl K, Hatt H. 2005. Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster*. *Nature Neuroscience* 8: 15-17.
- Ohler U, Liao GC, Niemann H, Rubin GM. 2002. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biology* 3: 1-12.
- Pitts RJ, Fox AN, Zwiebel LJ. 2004. A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector *Anopheles gambiae*. *Proceedings of the National Academy of Sciences USA* 101: 5058-5063.
- Robertson HM, Wanner KW. 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Research* 16: 1395-1403.
- Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA* 100: 14537-14542.
- Rutzler M, Zwiebel LJ. 2005. Molecular biology of insect olfaction: recent progress and conceptual models. *Journal of comparative physiology A, Neuroethology sensory, neural, behavioral physiology* 191: 777-790.
- Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, Yasukochi Y, Touhara K, Nishioka T. 2004. Identification and functional characterization of a sex pheromone receptor in the silkworm *Bombyx mori*. *Proceedings of the National Academy of Sciences USA* 101: 16653-16658.
- Scott K, Brady R, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R. 2001. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell* 104: 661-673.
- Suh GS, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ. 2004. A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* 431: 854-859.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.
- Vosshall LB. 2003. Diversity and expression of odorant receptors in *Drosophila*. In: Blomquist G, Vogt R, editors. *Insect Pheromone Biochemistry and Molecular Biology*, pp. 567-591. Elsevier Academic Press.
- Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* 96: 725-736.
- Vosshall LB, Wong AM, Axel R. 2000. An olfactory sensory map in the fly brain. *Cell* 102: 147-159.
- Wetzel CH, Behrendt HJ, Gisselmann G, Stortkuhl KF, Hovemann B, Hatt H. 2001. Functional expression and characterization of a *Drosophila* odorant receptor in a heterologous cell system. *Proceedings of the National Academy of Sciences USA* 98: 9377-9380.
- White JH, Wise A, Main MJ, Green A, Fraser NJ, Disney GH, Barnes AA, Emson P, Foord SM, Marshall FH. 1998. Heterodimerization is required for the formation of a functional GABA(B) receptor. *Nature* 396: 679-682.
- Wistrand M, Käll L, Sonnhämmer E. 2006. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Science* 15: 509-521.
- Xia Y, Zwiebel LJ. 2006. Identification and characterization of an odorant receptor from the West Nile virus mosquito, *Culex quinquefasciatus*. *Insect Biochemistry and Molecular Biology* 36: 169-176.
- Young L, Jernigan RL, Covell DG. 1994. A role for surface hydrophobicity in protein-protein recognition. *Protein Science* 3: 717-729.