

Original Article

Sequence-Based Prediction of Olfactory Receptor Responses

Shashank Chepurwar¹, Abhishek Gupta^{1,2}, Rafi Haddad³ and Nitin Gupta^{1,*}

¹Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India, ²Department of Chemistry, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India and ³The Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, 5290002, Israel

Correspondence to be sent to: Nitin Gupta, Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh 208016, India. e-mail: guptan@iitk.ac.in

Editorial Decision 23 August 2019.

Abstract

Computational prediction of how strongly an olfactory receptor (OR) responds to various odors can help in bridging the widening gap between the large number of receptors that have been sequenced and the small number of experiments measuring their responses. Previous efforts in this area have predicted the responses of a receptor to some odors, using the known responses of the same receptor to other odors. Here, we present a method to predict the responses of a receptor without any known responses by using available data about the responses of other conspecific receptors and their sequences. We applied this method to ORs in insects *Drosophila melanogaster* (both adult and larva) and *Anopheles gambiae* and to mouse and human ORs. We found the predictions to be in significant agreement with the experimental measurements. The method also provides clues about the response-determining positions within the receptor sequences.

Keywords: anopheles, *drosophila*, olfaction, olfactory, receptors

Introduction

Odor sensing begins with the binding of odor molecules to olfactory receptors (ORs) expressed on the membranes of olfactory sensory neurons. An organism often expresses a repertoire of tens or hundreds of types of receptors, which combinatorially can detect a very large number of odor stimuli (Malnic et al. 1999). The amino-acid sequences of these receptors—and consequently their tuning profiles to odors—vary from species to species in accordance with their environmental niches.

Identification of the odors that a given OR responds to, also known as deorphanization, is fundamental to understanding olfactory processing. In cases where the sensory neurons expressing a particular OR are known and easy to target with electrodes, the cognate odors can be identified using extracellular recordings from those neurons (Lu et al. 2007; Olsson and Hansson 2013; Li et al. 2018). An alternative approach is to express the OR in an easy-to-target cell

using heterologous expression systems—including cell lines, *Xenopus* oocytes, or the empty neuron system; these have been used in different species, including humans (Touhara et al. 1999; Mainland et al. 2015), mice (Oka et al. 2006; Saito et al. 2009), fruit flies (de Bruyne et al. 2001; Hallem and Carlson 2006; Lin and Potter 2015), mosquitoes (Carey et al. 2010; Wang et al. 2010), moths (de Fouchier et al. 2017), and tsetse flies (Chahda et al. 2019). However, such experiments need to be performed separately for each receptor and are time consuming; moreover, some receptors do not express well in the heterologous systems (Ronderos et al. 2014). High-throughput methods for identifying OR–odor interactions have also been developed recently, but these require elaborate experimental pipelines for each species (Jiang et al. 2015; Hu and Matsunami 2018; Jones et al. 2019). The high-throughput methods also generate many false positives, so it is recommended that the identified responses be further verified with targeted experiments (Nishizumi and Sakano 2015).

With lowering costs of sequencing, the sequences of ORs are rapidly becoming available for various organisms, including multiple species of fruit flies (*Drosophila* 12 Genomes Consortium 2007; Gardiner et al. 2008; Ramasamy et al. 2016), mosquitoes (Leal et al. 2013; Neafsey and Waterhouse 2015; Lombardo et al. 2017), bees (Elsik et al. 2016; Karpe et al. 2017), and tsetse flies (Attardo et al. 2019). However, experimental data about odor responses are available for relatively few ORs. Computational predictions can potentially bridge this gap. Indeed, computational methods have been developed to predict the response of an OR to new odors based on its known responses to other odors by analyzing the chemical structure (Schmucker et al. 2007), molecular volume (Sabeti and Seyed-Allaei 2016), or other physicochemical parameters of odor molecules (Haddad et al. 2008; Boyle et al. 2013; Gabler et al. 2013; Bushdid et al. 2018; Kepchia et al. 2019). However, these methods can only be used for ORs whose responses to some odors are known.

Here, we present a computational approach for predicting the responses of an OR even when no odor responses are available for that OR; our method instead uses the known responses of other conspecific ORs and the sequence similarities among them. We demonstrate the effectiveness of this approach by comparing our predictions with the experimentally measured responses in insects. Our method also provides insights into the response-determining positions in the receptor sequences.

Results

Lack of correlation between OR sequences and responses

As odor responses of an OR depend on its 3D structure, which in turn depends on the amino-acid sequence, we expected that the sequence similarity between a pair of ORs will relate to the similarity in their responses. To test this idea, we used large-scale data sets of odor responses of 24 ORs in *Drosophila melanogaster* (Hallem and

Carlson 2006) and 50 ORs in *Anopheles gambiae* (Carey et al. 2010) to sets of 110 odors; these odor sets were partially different between the 2 species but identical for all ORs within a species. For each pair of ORs within a species, we calculated the distance between their 110-length response vectors (see Materials and methods). We quantified amino-acid sequence similarity for each pair of ORs within a species using global sequence alignment. While we expected to see a negative correlation between sequence similarity and response distance, we surprisingly found no correlation at all (Figure 1): the Pearson's correlation coefficient was 0.0018 ($P = 0.97$, $N = 276$ pairs of receptors) in *D. melanogaster* and 0.026 ($P = 0.35$, $N = 1225$ pairs) in *A. gambiae*.

Identification of response-determining positions

The task of sequence-based predictions is made challenging by the very low level of conservation among the ORs: the percentage amino-acid identity between pairs of ORs was only $20.23 \pm 4.48\%$ in *D. melanogaster* and $20.60 \pm 8.17\%$ in *A. gambiae* (mean \pm standard deviation (SD) calculated over all pairs within a species). The lack of correlation between the sequences and the responses is partly because only a small fraction of residues in the sequence are involved in determining odor specificity, and the similarity of these residues is not represented well by the overall sequence similarity (Man et al. 2007). However, as there is very little structural information available for ORs, the positions of these response-determining residues are not known. We took an empirical approach to rank each position in the multiple sequence alignment of all receptor sequences according to its importance in determining the odor responses (see Materials and methods). Within a species, the ranking was done jointly over all ORs to avoid overfitting. Using this ranked list of positions for each species, we selected the top 20 positions that should help in predicting the odor responses of new ORs.

We first confirmed that the subsequences of amino acids formed by these 20 positions in each OR were indeed related to the responses.

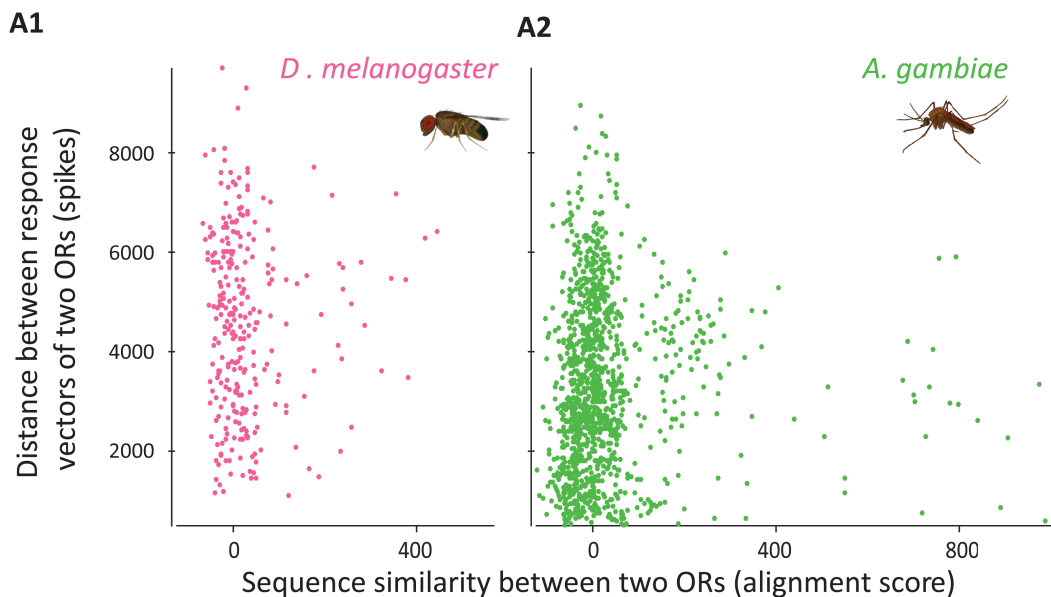


Figure 1. Lack of correlation between response distance and sequence similarity among pairs of ORs. Scatter plots of distances in response vectors versus the amino-acid sequence similarity (a larger value indicates more similarity), among pairs of ORs, reveal the lack of correlation in both *D. melanogaster* (A1) and *A. gambiae* (A2). Distance is calculated as the L1-distance between the two 110-length response vectors containing number of spikes for each of the 110 odors. Sequence similarity is measured as the Needleman–Wunsch alignment score. Each point corresponds to a pair of conspecific ORs ($N = 276$ pairs in *D. melanogaster* and $N = 1225$ pairs in *A. gambiae*).

We found the correlation between the response distance and sequence similarity (calculated using only the subsequences formed by the 20 positions) to be -0.56 ($P = 1.7 \times 10^{-24}$, $N = 276$ pairs; Figure 2A1) in *D. melanogaster* and -0.39 ($P = 1.7 \times 10^{-45}$, $N = 1225$ pairs; Figure 2A2) in *A. gambiae*. Compared to the result seen with the full sequence (Figure 1), these correlations were significantly negative. Similar negative correlations were observed even if we used only half of the odors for calculating the top 20 positions and the other half for calculating the response similarity (Supplementary Figure S1). If we varied the number of top-ranked positions used in the analysis, the correlation coefficients did not change for numbers between 10 and 40 but gradually reduced for larger numbers of positions (Figure 2B1,2), suggesting that lower-ranked residues are not important for the prediction of responses. Subsequently, we used the top 20 positions as the response-determining positions in our analysis (Figs. 2B1,2 show that our results would not change if we chose some other number between 50% and 200% of this number).

We then checked the robustness of the top 20 positions by reducing the number of receptors used for ranking the positions. When we removed 1 receptor (this was done by removing each receptor, one at a time, to get the averages and error bars), we found that on average ~ 16 of the 20 positions in *D. melanogaster* and ~ 19 of the 20 positions in *A. gambiae* were retained in the top 20. The number of retained positions dropped gradually as more receptors were left

out (again, all possible combinations were tested for each number). But even with half of the receptors removed, at least 8 of the original 20 positions remained in top 20 in both the species (Figure 2C1,2). This analysis indicated that the selected positions were robust to minor perturbations in the data sets.

Although the top 20 alignment positions were chosen independently in *D. melanogaster* and *A. gambiae*, we found that 3 of the top 20 positions were conserved between the species. This number was 3 times the number of conserved positions expected by chance ($20 \times 20/369 = 1.08$, where 369 is the number of alignment positions that were ranked in both the species), further increasing our confidence in the reliability of the identified positions.

Predicting responses using the identified positions

Having identified reasonable candidates for response-determining positions, common to all ORs within a species, we asked if it is possible to reliably predict the responses of a query OR to a panel of odors using only the sequence information about the amino acids present at the identified positions. We tested this idea using a simple and intuitive approach, in which the response of a query OR was estimated as the mean of the responses of known ORs, weighted by their similarity to the query OR at the response-determining positions using the BLOSUM62 amino-acid substitution matrix (see Materials and methods).

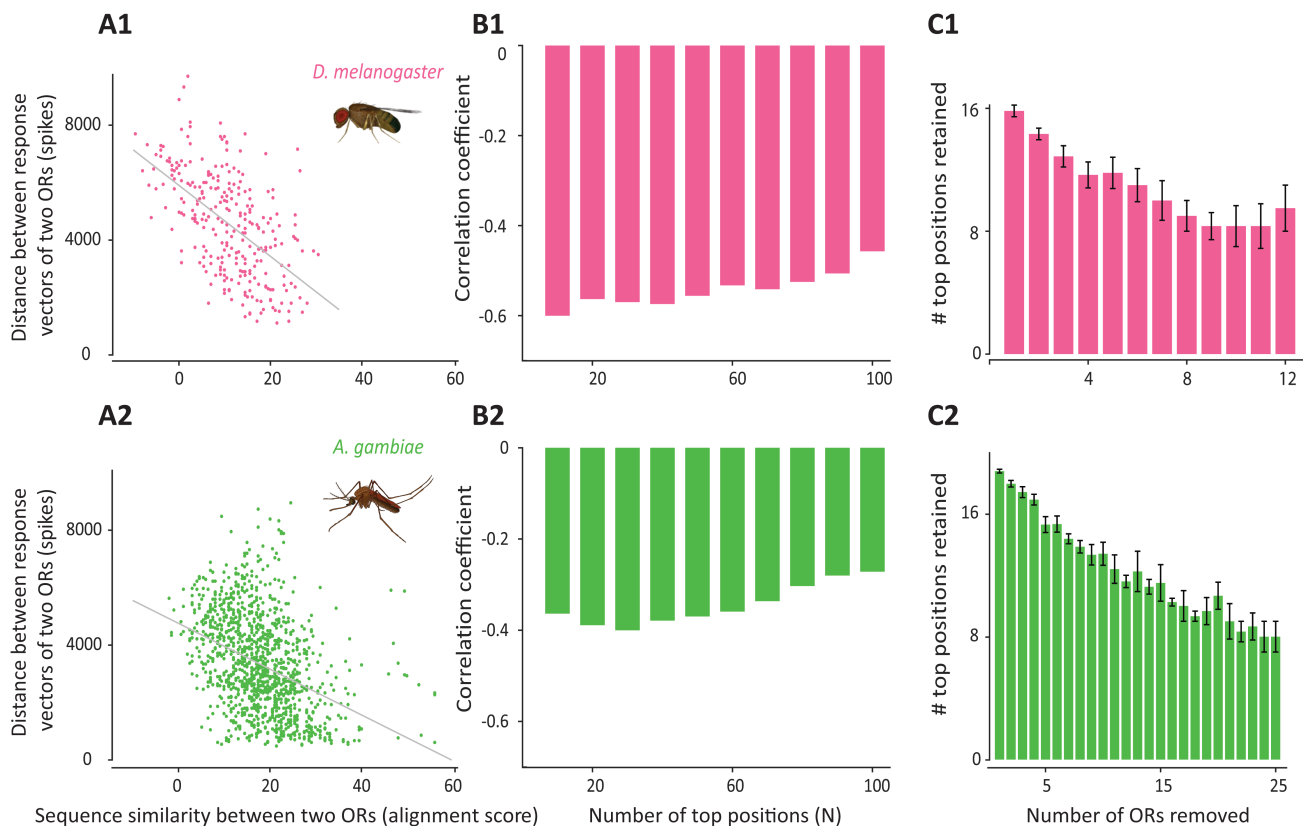


Figure 2. Analysis with response-determining positions. (A1, A2) Scatter plots of distances in 110-odor response vectors versus the sequence similarity measured at the top 20 positions, among pairs of ORs, show significant negative correlation in both *D. melanogaster* (A1) and *A. gambiae* (A2). Each point corresponds to a pair of conspecific ORs ($N = 276$ pairs in *D. melanogaster* and $N = 1225$ pairs in *A. gambiae*). (B1, B2) Bars indicate the negative correlation between the distance in response vectors and the sequence similarity measured using the top N positions, plotted as a function of N in the range of 10–100. Note that the correlation values are relatively stable in both the species in the range of 10–40 and decrease gradually in magnitude for larger values of N . (C1, C2) Bars indicate the number of top 20 positions in *D. melanogaster* (C1) or *A. gambiae* (C2) retained among the top 20 positions when the positions were recalculated after removing different numbers of receptors from the data set. For each number, all possible combinations of that many receptors were removed and the average number of positions preserved among the top 20 is reported. Error bars represent SEM over all combinations for each number.

We first applied this approach to predict the responses of the 24 and 50 ORs in the *D. melanogaster* and *A. gambiae* data sets, respectively. By turn, each OR was treated as a query OR and its response was predicted using the responses of the remaining ORs

in the same species (excluding the query OR). We found that the predicted response profiles (i.e., the 110-length vectors of odor responses) of the ORs were correlated with their actual response profiles (Figure 3A,B); the correlations were positive and statistically

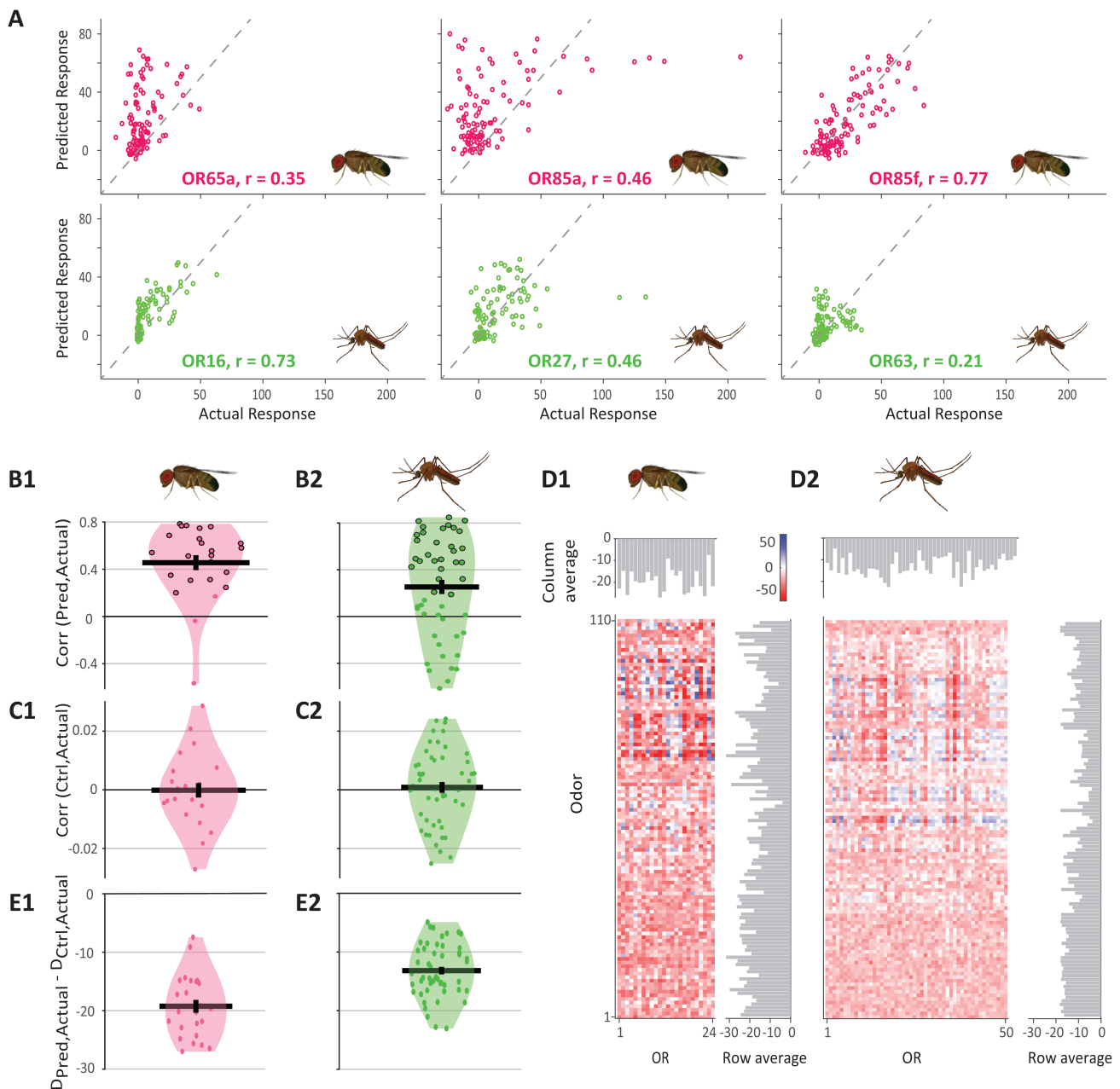


Figure 3. Reliability of responses predicted using the top 20 positions. **(A)** Scatter plots of the actual and the predicted odor responses for 3 representative ORs each from *D. melanogaster* and *A. gambiae*. The correlation coefficients between the 2 responses are indicated. The responses were spread around the $y = x$ line (dashed line, corresponding to ideal predictions). **(B1, B2)** Violin plots showing the correlation coefficient between the predicted and the actual response vectors of ORs in *D. melanogaster* (B1) or *A. gambiae* ($N = 50$) (B2). Each point corresponds to one OR ($N = 24$ ORs in *D. melanogaster* and $N = 50$ ORs in *A. gambiae*). Note that the correlations are mostly positive. Statistically significant positive correlations (with P less than 0.05) are shown with black outlines. **(C1, C2)** Violin plots showing the correlation coefficient between the control predictions and the actual response vectors of ORs in *D. melanogaster* (C1) or *A. gambiae* ($N = 50$; C2). Note that the correlations are close to zero in most cases. **(D1, D2)** Color maps showing the relative prediction error for each OR–odor pair, for the set of 110 odors, and all 24 ORs in *D. melanogaster* (D1) or all 50 ORs in *A. gambiae* (D2). The relative prediction error is measured as $D_{Pred,Actual} - D_{Ctrl,Actual}$, where $D_{Pred,Actual}$ is the absolute distance between the predicted and actual response and $D_{Ctrl,Actual}$ is the absolute distance between the control prediction and the actual response. The abundance of negative values (red shades) indicates that the predictions are usually better than the control. Column averages and row averages of the matrix values are shown on top and right, respectively. **(E1, E2)** Violin plots shows the average (over odors) of $D_{Pred,Actual} - D_{Ctrl,Actual}$ for each of the 24 ORs in *D. melanogaster* (E1) or 50 ORs in *A. gambiae* (E2). In violin plots, horizontal line and the error bar indicate the mean and the SEM, respectively, within a species.

significant in 21 of the 24 ORs in *D. melanogaster* and 29 of the 50 ORs in *A. gambiae*. Mean of these correlation values was significantly greater than 0: 0.46 ($P = 2.53 \times 10^{-7}$, t -test; $P = 1.63 \times 10^{-5}$, sign-rank test; $N = 24$ ORs) in *D. melanogaster* and 0.25 ($P = 1.18 \times 10^{-4}$, t -test; $P = 1.19 \times 10^{-4}$, sign-rank test; $N = 50$ ORs) in *A. gambiae*. In contrast, control predictions obtained by shuffling the response matrices (see Materials and methods) were not correlated with the actual responses (Figure 3C): mean correlation values were 0.0002 ($P = 0.95$, t -test; $N = 24$ ORs) in *D. melanogaster* and 0.0008 ($P = 0.65$, t -test; $N = 50$ ORs) in *A. gambiae*.

The correlation coefficient between the actual and the predicted odor response vectors tells whether these responses have the same relative magnitude across different odors but does not tell how close the 2 responses are in absolute terms (a predicted response that is several times larger than the actual response for every odor will have a perfect correlation). We, therefore, also used a distance-based metric and checked whether our predictions were closer to the actual responses than the control predictions were. We define $D_{\text{Ctrl, Actual}}$ as the absolute difference between the predicted response and the actual response of an OR to an odor and $D_{\text{Ctrl, Actual}}$ as the absolute difference between the control prediction and the actual response. Thus, these 2 terms indicate the error in the actual and the control predictions. If our predicted response is better (closer to the actual response) than the control prediction, the value of $D_{\text{Pred, Actual}} - D_{\text{Ctrl, Actual}}$ should be negative. The heat maps in Figure 3D show the comparisons for all the predictions in *D. melanogaster* and *A. gambiae* and reveal a very high abundance of negative values. In *D. melanogaster*, the average $D_{\text{Pred, Actual}}$ of an OR (28.60) was smaller than average $D_{\text{Ctrl, Actual}}$ (47.84) by 19.24 spikes ($P = 7.30 \times 10^{-15}$, paired t -test; $N = 24$ ORs; Figure 3E1), an improvement of more than 40%. In *A. gambiae*, the average $D_{\text{Pred, Actual}}$ (20.11) was smaller than average $D_{\text{Ctrl, Actual}}$ (33.29) by 13.18 spikes ($P = 8.46 \times 10^{-26}$, paired t -test; $N = 50$ ORs; Figure 3E2), again an improvement of about 40%. Thus, our predicted responses using the sequence similarity at the response-determining positions were significantly closer to the actual responses than random predictions were. These improvements were substantial, considering the absolute values of odor responses in *D. melanogaster* (mean \pm SD = 32.43 ± 49.29 spikes, $N = 24 \times 110$ OR–odor combinations) and *A. gambiae* (mean \pm SD = 20.31 ± 37.35 spikes, $N = 50 \times 110$).

Test on an independent data set

Next, we collected the responses of 26 other ORs in *D. melanogaster* from 10 different studies (de Bruyne et al. 1999, 2001, 2010; Dobritsa et al. 2003; Goldman et al. 2005; Kreher et al. 2005, 2008; Marshall et al. 2010; Ronderos et al. 2014; Dweck et al. 2015) deposited in the Database of Odorant Responses (DoOR) (Münch and Galizia 2016). Each of these ORs had responses available for some of the 110 odors used previously; overall, 68 of the previously used odors were represented in the new data set. These 26 ORs were entirely different from the 24 ORs used previously in selecting the response-determining positions and, therefore, provided an independent test for our approach. Using the set of top 20 response-determining positions as selected earlier from the original *D. melanogaster* data set, we found that the responses predicted for the 26 novel ORs were significantly better with our approach compared to the control predictions (Figure 4A,B): the average $D_{\text{Pred, Actual}}$ of an OR (37.01) was smaller than average $D_{\text{Ctrl, Actual}}$ (51.36) by 14.35 spikes ($P = 1.13 \times 10^{-7}$, paired t -test; $P = 5.96 \times 10^{-8}$, sign-rank test; $N = 26$ ORs), an improvement of about 28%. Overall, among the 506 OR–odor responses predicted (nongray squares in Figure 4A), in 380 (75.1%)

cases, the predictions were better than the control predictions. These results show that once the response-determining positions have been identified using known responses of some ORs in a species, the responses of other conspecific ORs can be predicted reasonably using only their sequences.

While predicting the responses of the new ORs above, we had relied on the known responses of 24 ORs. Can similar predictions be made if responses of fewer ORs are available? To check this, we sampled subsets of ORs from the original data set of 24 ORs and used these samples to select response-determining positions and make the response predictions. The sizes of these samples varied from 3 to 21 in intervals of 3 and, for each size, the sampling was performed 10 times. We found that the difference between the predicted and the actual response ($D_{\text{Pred, Actual}}$) decreased gradually as the number of sampled ORs increased but saturated for 15 or more ORs (Figure 4C). A similar trend was observed when we checked how the number of sampled ORs affected the fraction of predictions that were better than control (Figure 4D). These results suggest that a relatively small set of ORs with known responses may be enough to make reasonable predictions for new ORs.

Predicting responses in larval *Drosophila*

Although the OR responses in larval *D. melanogaster* differ from the responses in adult *Drosophila* (Kreher et al. 2008), the ORs themselves are the same at the molecular level. Therefore, we checked if the top positions identified in adult *Drosophila* can be used to predict the larval responses. We tested this idea on a data set of 21 ORs and 26 odors (each available at 2 different concentrations, 10^{-2} and 10^{-4}) from larval *Drosophila* (Kreher et al. 2008). For each concentration, we found that the predictions were significantly better than control predictions (Supplementary Figure S2): at 10^{-2} concentration, the average $D_{\text{Pred, Actual}}$ (38.01) was smaller than average $D_{\text{Ctrl, Actual}}$ (54.72) by 16.71 spikes ($P = 6.32 \times 10^{-12}$, paired t -test; $N = 21$ ORs); at 10^{-4} concentration, the average $D_{\text{Pred, Actual}}$ (12.38) was smaller than average $D_{\text{Ctrl, Actual}}$ (16.7) by 4.32 spikes ($P = 1.1 \times 10^{-8}$, paired t -test; $N = 21$ ORs).

Properties of the response-determining positions

Our results show that the identified subset of sequence positions can allow reliable prediction of the neural responses for new ORs. To check where these top 20 positions are located within the protein structure, we visualized these positions in the secondary structure of 2 representative ORs in *D. melanogaster* and *A. gambiae* using the Protter tool (Omasits et al. 2014). These residues were not limited to any one region but were instead found in various parts of the protein (Figure 5). It is important to note that our analysis does not imply a causal or a mechanistic role for each of these residues in interacting with the odor molecules but only suggests that these residues may be involved in some aspect of determination of odor responses, perhaps indirectly. Notably, the region with the highest density of the top 20 positions was the seventh transmembrane helix, including 4 positions in *D. melanogaster* and 5 in *A. gambiae* (Figure 5; similar numbers were seen in other receptors). Moreover, of the 3 response-determining positions that were conserved between the 2 species, 2 were located in the seventh transmembrane helix. These observations are in agreement with previous studies, suggesting that the seventh transmembrane helix plays an important role in determining odor specificity (Hughes et al. 2014; Ray et al. 2014).

Next, we collected data from various experimental studies that have mutated individual residues in various ORs and have checked their effect on odor responses. We collected a total of 47 such mutated

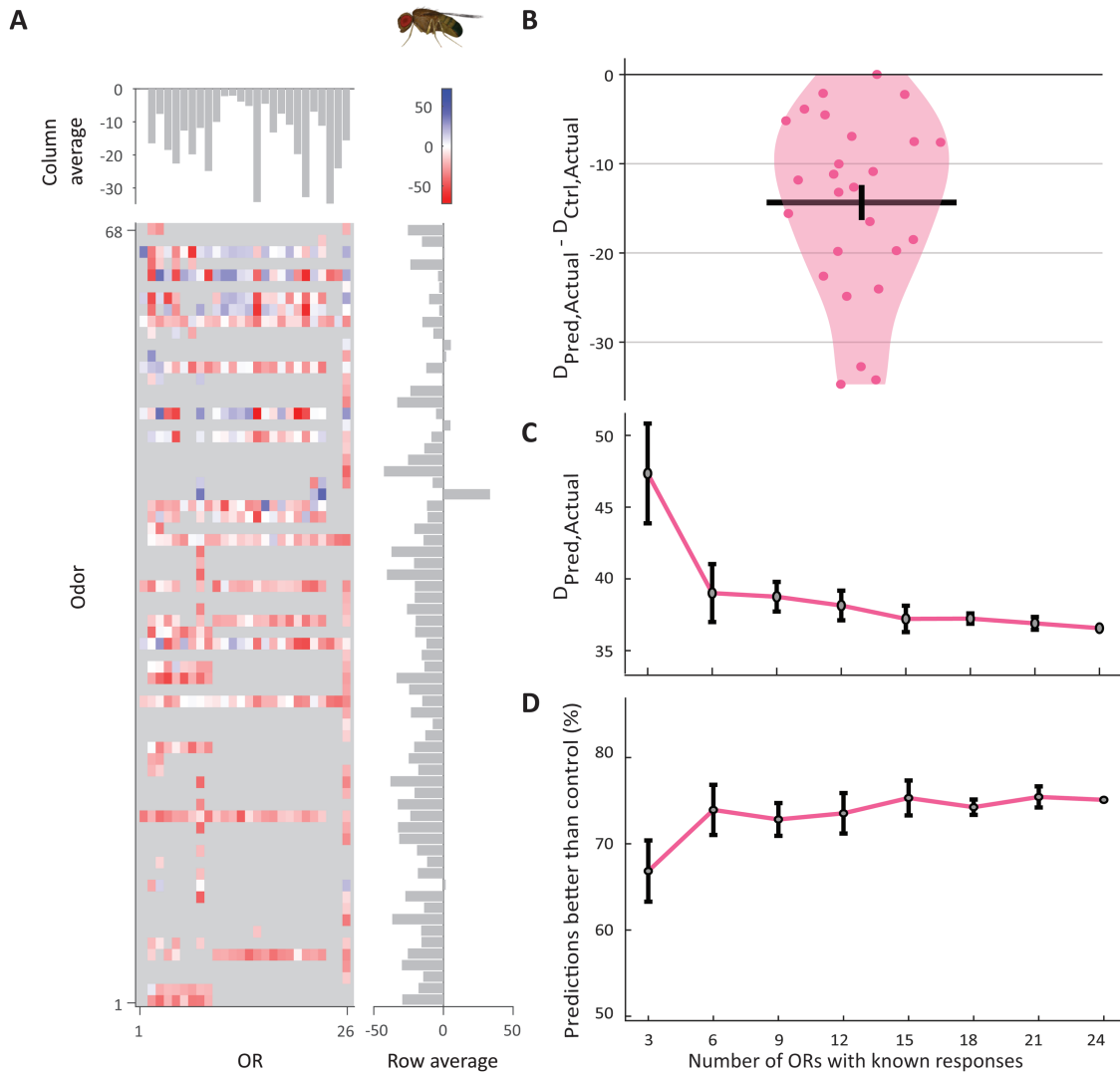


Figure 4. Response prediction on an independent set of ORs. **(A)** Color map showing the relative prediction error (measured as $D_{\text{Pred,Actual}} - D_{\text{Ctrl,Actual}}$) for each OR–odor pair for the set of 68 odors and 26 novel ORs. Pairs whose values were not available are shown in gray. Note the high frequency of negative values (red) and very low frequency of positive values (blue), indicating that the predictions were closer to the actual value than the control were in most cases. **(B)** Violin plots shows the average (over odors) of $D_{\text{Pred,Actual}} - D_{\text{Ctrl,Actual}}$ for each of the 26 ORs. Horizontal line and the error bar indicate the mean and the SEM, respectively. **(C)** $D_{\text{Pred,Actual}}$ is plotted as a function of the number of ORs with known responses used in making the predictions. These different numbers of ORs were sampled from the original set of 24 ORs with known responses. For each number, the sampling was performed 10 times; for each sampling, selection of response-determining positions and prediction of responses was performed. **(D)** For each number of sampled ORs, we calculated the fraction of the 506 OR–odor pairs (nongray values in panel A), for which the predicted responses were closer to the actual values than the control predictions were. Both methods of generating control predictions (see Materials and methods) were compared. In panels C and D, error bars indicate SEM over the 10 samplings for each number from 3 to 21; for 24, all available ORs were used.

positions from 5 studies (Nichols and Luetje 2010; Nakagawa et al. 2012; Xu and Leal 2013; Hughes et al. 2014; Ray et al. 2014) and compared them with our identified set of response-determining positions. In addition to *D. melanogaster* and *A. gambiae*, these studies also included data from *Bombyx mori* and *Culex quinquefasciatus*. Sequences from these other species were added to our existing multiple sequence alignment (see Materials and methods), so that the mutated positions in all these species can be compared with our alignment positions.

Given 20 response-determining positions in *D. melanogaster* and *A. gambiae* each, and with 3 being common between the 2 species, 37 unique positions in the multiple sequence alignment were marked as the top positions out of a total of 817. Thus, each of the 47 mutated positions in the curated experimental set had a 37/817

probability of matching a top position, and the expected number of matches by chance = $47 \times 37/817 = 2.13$. So, while only about 2 of the 47 mutated positions were expected to be among the top positions by chance, we found as many as 9 of them to be among the top positions ($P = 2.28 \times 10^{-4}$; binomial test), providing experimental support to our response-determining positions. These 9 experimentally studied mutations, which map to the top positions in our alignment, are listed in Table 1 (further, Supplementary Table S1 shows all positions). Among these, residue 146 in OR85b of *D. melanogaster* is one of the residues expected to play a crucial role in the activation of the odorant (Nichols and Luetje 2010). Residue 109 mutant in OR1 of *B. mori* was found to change the reversal potential and rectification index of the BmOR1–Orco complex (Nakagawa et al. 2012). Point mutation of residues 165 and 194 in BmorOR1 of *B. mori* led

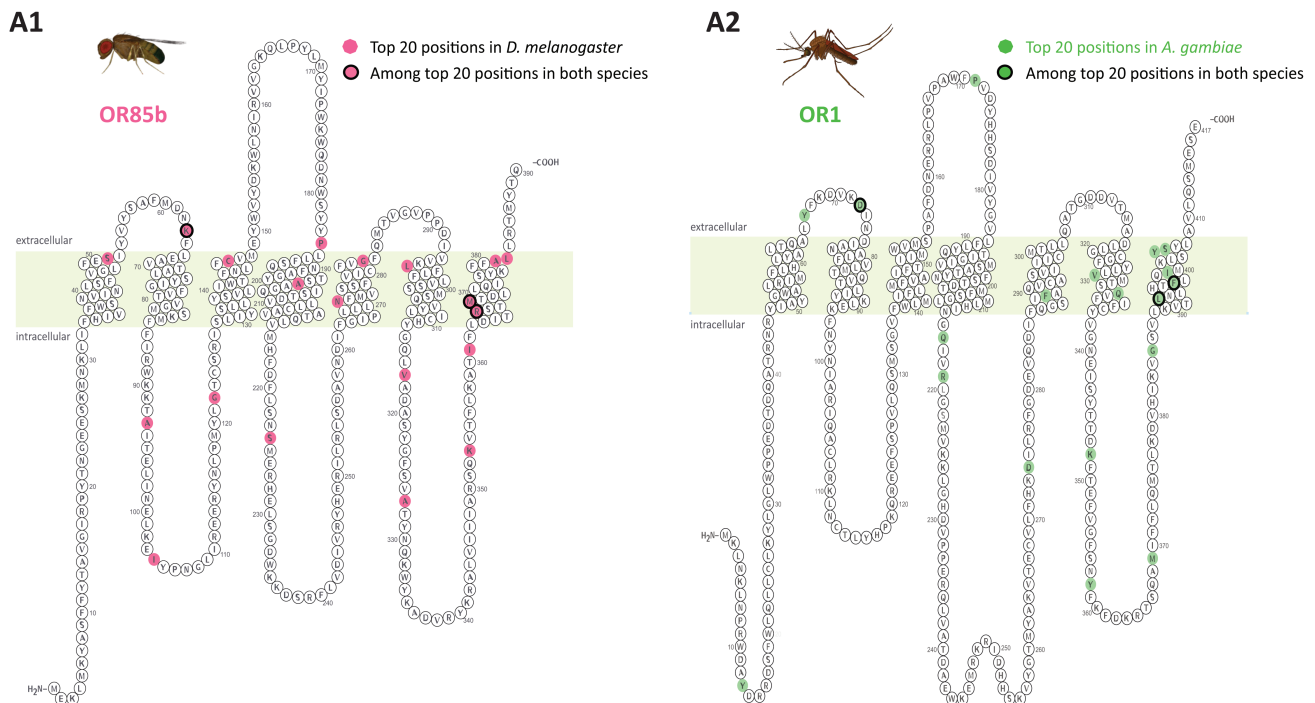


Figure 5. Localization of the top 20 response-determining positions. Protter plots showing the localization of the top 20 response-determining residues in representative ORs: OR85b of *D. melanogaster* (A1) and OR1 of *A. gambiae* (A2). Residues in the top 20 positions in the corresponding species are shown with shaded backgrounds, while the residues that are among the top 20 positions in both the species are shown with a black outline. One of the top 20 positions in *A. gambiae* mapped to a gap in OR1 in the multiple sequence alignment, so only 19 positions are shaded.

Table 1. Positions identified in previous mutagenesis studies that match the top positions identified in our analysis.

Study	Mutated receptor	Mutated residue	Position in multiple sequence alignment	Rank among top positions	Topological position
Hughes et al. (2014)	<i>A. gambiae</i> OR15	F369L	783	2 in <i>D. melanogaster</i> and 17 in <i>A. gambiae</i>	TM 7
Nakagawa et al. (2012)	<i>B. mori</i> OR1	Y109A	234	14 in <i>D. melanogaster</i>	TM 3
Nichols and Luetje (2010)	<i>D. melanogaster</i> OR85b	C146S	335	20 in <i>D. melanogaster</i>	TM 3
Ray et al. (2014)	<i>A. gambiae</i> OR1	Y405H	791	6 in <i>A. gambiae</i>	TM 7
Ray et al. (2014)	<i>A. gambiae</i> OR1	S406A	792	20 in <i>A. gambiae</i>	TM 7
Xu and Leal (2013)	<i>B. mori</i> OR1	P165A	335	20 in <i>D. melanogaster</i>	TM 3
Xu and Leal (2013)	<i>B. mori</i> OR1	P194A	398	9 in <i>A. gambiae</i>	EL 2
Xu and Leal (2013)	<i>A. gambiae</i> OR10	P154A	398	9 in <i>A. gambiae</i>	EL 2
Xu and Leal (2013)	<i>C. quinquefasciatus</i> OR10	P157A	398	9 in <i>A. gambiae</i>	EL 2

Of the 47 mutations that we collected from the mutagenesis studies, the 9 that matched the top positions in *D. melanogaster* or *A. gambiae* are listed. The complete set of the 47 mutations is shown in [Supplementary Table S1](#). EL, extracellular loop; TM, transmembrane helix.

to significant reduction in response to bombykol pheromone (Xu and Leal 2013). Similarly, odor responses in *A. gambiae* were found to change with mutation of residue 369 in OR15 (Hughes et al. 2014) and residues 405 and 406 in OR1 (Ray et al. 2014). Among the remaining 38 out of the 47 mutations in the curated set, 10 and 8 were found to be within a distance of 1 and 2 residues, respectively, from a top column. Thus, in total, 27 of the 47 experimentally confirmed important positions matched or were adjacent to our computationally predicted response-determining positions.

Prediction of vertebrate olfactory responses

As our method does not make any assumption about the composition of the receptor sequences, it should be applicable to any family of chemosensory receptors. We next sought to check its utility in

predicting the responses of the vertebrate ORs. Saito et al. (2009) have provided a data set of EC50 values of 52 mouse and 10 human ORs that responded to at least 1 of the 63 odors tested. We extracted a binary response matrix (1 for response, 0 for no response) from this data set: among the mouse responses ($52 \times 63 = 3276$ values), this matrix included 262 (approximately 8%) 1s and rest were 0s. Using the multiple sequence alignment provided in the same study (Saito et al. 2009) and mouse data, we calculated the top 20 response-determining positions. We then calculated the response of each mouse receptor using the weighted average of the responses of the remaining receptors as done for insect receptors (see Materials and methods); for comparison with actual responses, this value was also converted to a binary response using a threshold, set to keep the total fraction of 1s in the predicted responses also at 8%.

The predictions, thus, obtained had a false alarm rate of about 7% and a hit rate of about 21%, resulting in sensitivity index $d' = 0.68$ ($P < 0.001$, estimated by comparing the d' value with those obtained for 1000 shuffled control predictions). We next asked if human OR responses could also be predicted using the mouse training data. Using the top positions, the responses, and the threshold identified from the mouse data, we predicted the binary responses of the 10 human ORs. Among the 10×63 values in the original human data set, 78 (~12%) were 1s. Our prediction yielded 581 0s and 49 1s, with a false alarm rate of about 7% and a hit rate of about 13% ($d' = 0.34$, $P = 0.047$ based on 1000 shuffled predictions). In summary, these results show that our method can also be used to predict vertebrate OR responses with better than chance accuracy.

Discussion

Sequence-based approaches to response prediction require a reasonably high homology among the receptors. In vertebrates, receptors with less than 40% sequence identity were found to have little overlap in their responses (Li et al. 2015). The insects ORs presented a particularly difficult test case as they have only about 20% homology among conspecific ORs; indeed, we found that there was no correlation between the overall sequence similarity and the response similarity among the ORs. We were able to overcome this difficulty by focusing on subsets of positions that may be important for determining odor responses. Our approach was successful in predicting the responses for not only the set of fly and mosquito receptors whose data was used while developing the method but also for a completely novel set of receptors whose responses were taken from 10 independent experimental data sets. Further, we found that the predictions require only about 15 ORs with known responses (Figure 4). This computational approach can be particularly useful for receptors expressed in neurons that are difficult to access in electrophysiology experiments or for receptors that do not express well in heterologous expression systems (Ronderos et al. 2014). Finally, we also showed that our method can also be used for vertebrate receptors.

Our method allows prediction of responses of novel ORs for already studied odors. Previously developed methods allow prediction of responses of already studied ORs to novel odors (Schmucker et al. 2007; Haddad et al. 2008; Boyle et al. 2013; Gabler et al. 2013; Saberi and Seyed-Allaei 2016; Bushdid et al. 2018; Kepchia et al. 2019). These 2 approaches are complementary and can be combined in future work to enable the prediction of the responses of novel ORs to novel odors, which would help with large-scale deorphanization. While applying these approaches, care must be taken regarding the concentrations of the odors in the training data as changes in odor concentration can affect neural responses (Hallem and Carlson 2006; Olsen et al. 2010) and behavior (Wright et al. 2005; Yarali et al. 2009). Another caveat is that these approaches are currently limited to predicting the response magnitudes and ignore the temporal patterns of spikes observed in the OR neurons (Spors et al. 2006; Raman et al. 2010; Montague et al. 2011; Su et al. 2011; Grillet et al. 2016; Egea-Weiss et al. 2018).

Unlike vertebrate ORs which are G-protein coupled receptors, insect ORs are heteromeric ligand-gated ion channels (Wicher et al. 2008), composed of a specific OR and a conserved coreceptor Orco (Larsson et al. 2004). The structure of any insect receptor was not known until very recently when the structure of Orco from fig wasp was studied using Cryo-EM (Butterwick et al. 2018). The approach of predicting responses using the 3D modeling of ORs, which has

seen some success in vertebrates (Bavan et al. 2014), is unlikely to be very productive in insects in the near future given the very low sequence similarity among the ORs and with other proteins and the paucity of available structures. Therefore, sequence-based computational methods will be important for receptors whose structures and responses have not been determined experimentally. The sequence-based predictions are expected to fare better when the responses being predicted are in the same range as the responses in the training data set but may not fare as well in predicting uniquely strong ligands for a novel receptor (for the latter goal, 3D modeling may be necessary). Our computational predictions can be used to identify the candidate ligands to be tested experimentally. The method can also be extended to other types of chemosensory receptors, such as the ionotropic receptors and the gustatory receptors.

We have used a simple algorithm in our approach, where we first identify the important positions in the sequences and then use the similarity at those positions to make the predictions. It may be possible to achieve good predictions without explicitly identifying the important positions and instead using a machine learning-based approach that uses all the residues and learns a nonlinear mapping between the sequences and the responses. However, we preferred the simpler deterministic approach over the blackbox approaches typically used in machine learning because the deterministic approach provides a transparent rationale for and information about the sequence positions that are used for making the predictions. Expectedly, a majority of the identified positions were located in the extracellular loops or transmembrane helices (particularly the seventh transmembrane, which has been experimentally shown to be important for determining responses), but some were also found in intracellular loops. A previous study has shown that mutation of intracellular residues can affect the electrophysiological responses of neurons (Nakagawa et al. 2012). We emphasize that our approach only looks for correlations between the residues and responses, and it is likely that some of the identified positions, particularly those found in intracellular regions, affect the odor responses indirectly rather than through direct interactions with the odor molecules. By comparing the 47 positions previously found to be important in mutagenesis studies in various organisms (Nichols and Luetje 2010; Nakagawa et al. 2012; Xu and Leal 2013; Hughes et al. 2014; Ray et al. 2014), we found that 9 of them (a statistically significant fraction more than 4 times larger than that expected by chance) mapped to the top positions identified in our analysis (Table 1). This shows that our method is using reliable positions for making the predictions. The other top positions that have not been experimentally studied so far are good candidates for future mutagenesis studies and could help in understanding the mechanisms of odor-OR interactions. As more experimental evidence accumulates, the response predictions can be further improved by giving more weightage to the experimentally verified positions among the top positions.

Materials and methods

Response data sets

Responses of 24 *D. melanogaster* ORs were obtained from a previous study (Hallem and Carlson 2006) and responses of 50 *A. gambiae* ORs were obtained from another study (Carey et al. 2010) for a set of 110 odors. The response values indicate the odor-elicited spiking response of a neuron expressing the receptor minus the response for the diluent or the spontaneous firing rate. To serve as an independent data set, responses of 26 other *D. melanogaster* ORs were obtained from 10 studies (de Bruyne et al. 1999, 2001, 2010; Dobritsa et al.

2003; Goldman et al. 2005; Kreher et al. 2005, 2008; Marshall et al. 2010; Ronderos et al. 2014; Dweck et al. 2015) deposited in the DoOR (Münch and Galizia 2016) for 68 odors that were common with the earlier set of 110 odors. If the response of a particular OR to an odor was reported in multiple studies, we used the average of those values as the actual response.

Sequence analysis

The OR sequences for *D. melanogaster* and *A. gambiae* were obtained from the Database of Olfactory Receptors (Nagarathnam et al. 2014 and the Hill et al. 2002 study, respectively). The sequence similarity between any 2 receptor sequences was determined using the Needleman–Wunsch Algorithm using the BLOSUM62 scoring matrix (Needleman and Wunsch 1970; Henikoff and Henikoff 1992). Multiple sequence alignment was performed for a total of 140 (62 in *D. melanogaster* and 78 in *A. gambiae*) OR sequences using MAFFT multiple sequence alignment program, and additional sequences from other species (for validation) were added individually using the “-add” option (Katoh and Frith 2012).

The OCTOPUS tool (Viklund and Elofsson 2008) was used to predict the membrane topology of the receptors and the Protter tool was used to visualize the 2D topology of the receptors (Omasits et al. 2014).

Ranking the response-determining positions

We identified the response-determining sequence positions for *D. melanogaster* and *A. gambiae* using the following approach, which quantifies at each position how well the differences in amino acids across the receptors correlate with the differences in their responses. These positions corresponded to columns in the multiple sequence alignment and were same for all the receptors within a species (allowing different positions for each receptor can provide better correlations for individual receptors due to overfitting but will have poor generalizability for other receptors).

From the total of 817 alignment columns in the multiple sequence alignment, we first eliminated the columns which were either fully conserved or had gaps in more than half of the sequences, leaving 378 columns to be ranked in *D. melanogaster* and 392 columns to be ranked in *A. gambiae*, of which 369 were common to both species (see Supplementary Table S1). We denote an element of the alignment by $A_{r,c}$ where c indicates the alignment column and $r \in [1, 24]$ or $[1, 50]$ in *D. melanogaster* and *A. gambiae*, respectively, indicates the row corresponding to a receptor with known responses. The following steps were carried out separately for each species. To calculate the rank for each of the columns, we first calculated a sequence similarity vector (SS_c) for each column, each element ($SS_{(r_1,r_2),c}$) of which contained the pairwise amino acid similarity scores, calculated using BLOSUM62 matrix, between all pairs of receptors (r_1, r_2) that did not have a gap in the alignment column c :

$$SS_{(r_1,r_2),c} = \text{BLOSUM62}(A_{r_1,c}, A_{r_2,c}) \forall (r_1, r_2), \text{ where } A_{r_1,c} \neq \text{gap and } A_{r_2,c} \neq \text{gap}.$$

Similarly, we computed a response distance vector (DR_c^o) for the same pairs of ORs in the column, each element ($DR_{(r_1,r_2),c}^o$) of which contained the absolute difference between the responses of receptors r_1 and r_2 to an odor o :

$$DR_{(r_1,r_2),c}^o = |N_{r_1}^o - N_{r_2}^o| \forall (r_1, r_2), \text{ where } A_{r_1,c} \neq \text{gap and } A_{r_2,c} \neq \text{gap},$$

where N_r^o denotes the neural response of receptor r to odor o . We used the L1-norm instead of the Euclidean distance to avoid giving

extra weightage to larger values. Note that SS_c and DR_c^o are both vectors of the same length; this length is equal to the number of all possible pairs of receptors that did not have a gap in column c . We then calculated the Pearson's correlation coefficient (R_c^o) between the sequence similarity (SS_c) and response difference (DR_c^o) vectors for each column c and each odor o . The P -value of this correlation, denoted by P_c^o , was also noted. A negative value of R_c^o indicates that the receptors that have high sequence similarity at the alignment position c also have high response similarity (small distance) for odor o . Finally, we calculated the total score of a column as

$$S_c = - \sum_{\substack{o=1, \\ R_c^o < 0}}^{110} \log P_c^o$$

Use of P -values allowed us to give more weightage to highly significant correlations. Columns with high S_c scores were used subsequently as the response-determining positions.

Response prediction

We first generated the subsequences using the amino acids at the response-determining positions; this was done for the receptors with known responses and also for the receptor whose response is being predicted (query). Subsequence similarity ($SSS_{r, \text{query}}$) between the query receptor and each of the known receptors was calculated by taking the average of BLOSUM62 score for all positions in the 2 subsequences (excluding positions with gaps). The similarity values were then linearly scaled (denoted by $\overline{SSS}_{r, \text{query}}$) to a range of 0–100 to avoid negative weights in the subsequent steps:

$$\overline{SSS}_{r, \text{query}} = \frac{SSS_{r, \text{query}} - \min_i SSS_{r_i, \text{query}}}{\max_i SSS_{r_i, \text{query}} - \min_i SSS_{r_i, \text{query}}} \times 100$$

The predicted response (N_{query}^o) of the query receptor to an odor o was calculated as a weighted average of the known responses:

$$N_{\text{query}}^o = \frac{\sum_{j=1}^n (\overline{SSS}_{r_j, \text{query}} \times N_{r_j}^o)}{\sum_{j=1}^n \overline{SSS}_{r_j, \text{query}}}$$

where n is the number of receptors with known responses and $N_{r_j}^o$ denotes the neural response of receptor r_j to odor o .

Control predictions

Control predictions were obtained by shuffling the OR–odor response matrix and taking the shuffled response as the control prediction for each OR–odor combination (this operation allowed generation of random predictions while maintaining the overall statistical properties of the response values). The shuffling was not limited to shifting of rows or columns but included the shuffling of all elements in the matrix. The shuffling was performed 50 times independently, and the results (correlations or $D_{\text{Ctrl, Actual}}$ values) were averaged over the 50 shufflings.

Statistical analysis and code availability

All analyses were performed in MATLAB. A modified version of Gramm plotting toolbox (Morel 2018) was used to draw the plots. We used Pearson correlation to measure the correlations and t -tests and Wilcoxon signed-rank tests to compare the means. All tests were 2-tailed. The number of sample points is shown for each test in the results. The code developed in this study can be accessed from https://github.com/neuralsystems/OR_response_prediction.

Supplementary material

Supplementary data are available at *Chemical Senses* online.

Supplementary Table S1: Complete multiple sequence alignment of all the ORs.

The table shows the multiple sequence alignment of all the ORs that have been used in this study. Each row corresponds to one OR. The first 2 rows indicate with colored background the positions that were ranked in the 2 species; among these the ranks of the top 20 positions are mentioned. The 47 residues that have been identified as important in previous mutagenesis studies are marked with yellow color in the matrix.

Acknowledgement

We thank members of the N.G. laboratory for helpful comments.

Funding

This work was supported by the (IA/15/2/502091 to N.G.) and the ISF-UGC joint research program, in which R.H. was supported by the Israel Science Foundation (2307/15) and N.G. was supported by the University Grants Commission (6–11/2016[IC]).

Declaration of interests

The authors declare no competing financial interests.

Author contributions

N.G. conceived the study. S.C., A.G., R.H., and N.G. developed the methodology. S.C. and A.G. developed the software and performed analysis. N.G. and R.H. secured funding. S.C., A.G., R.H., and N.G. wrote the manuscript.

References

- Attardo GM, Abd-Alla AMM, Acosta-Serrano A, Allen JE, Bateta R, Benoit J, Bourtzis K, Caers JI, Caljon G, Christensen MB, et al. 2019. The glossina genome cluster: comparative genomic analysis of the vectors of African trypanosomes. *BioRxiv* 531749.
- Bavan S, Sherman B, Luetje CW, Abaffy T. 2014. Discovery of novel ligands for mouse olfactory receptor MOR42-3 using an in silico screening approach and in vitro validation. *PLoS One*. 9:e92064.
- Boyle SM, McInally S, Ray A. 2013. Expanding the olfactory code by in silico decoding of odor-receptor chemical space. *eLife*. 2:e01120.
- Bushdid C, de March CA, Fiorucci S, Matsunami H, Golebiowski J. 2018. Agonists of G-protein-coupled odorant receptors are predicted from chemical features. *J Phys Chem Lett*. 9:2235–2240.
- Butterwick JA, Del Marmol J, Kim KH, Kahlson MA, Rogow JA, Walz T, Ruta V. 2018. Cryo-EM structure of the insect olfactory receptor Orco. *Nature*. 560:447–452.
- Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. 2010. Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature*. 464:66–71.
- Chahda JS, Soni N, Sun JS, Ebrahim SAM, Weiss BL, Carlson JR. 2019. The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLoS Genet*. 15:e1008005.
- de Bruyne M, Clyne PJ, Carlson JR. 1999. Odor coding in a model olfactory organ: the *Drosophila* maxillary palp. *J Neurosci*. 19:4520–4532.
- de Bruyne M, Foster K, Carlson JR. 2001. Odor coding in the *Drosophila* antenna. *Neuron*. 30:537–552.
- de Bruyne M, Smart R, Zammit E, Warr CG. 2010. Functional and molecular evolution of olfactory neurons and receptors for aliphatic esters across the *Drosophila* genus. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 196:97–109.
- de Fouchier A, Walker WB 3rd, Montagné N, Steiner C, Binyameen M, Schlyter F, Chertemps T, Maria A, François MC, Monsemper C, et al. 2017. Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nat Commun*. 8:15709.
- Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR. 2003. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron*. 37:827–841.
- Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 450:203–218.
- Dweck HK, Ebrahim SA, Farhan A, Hansson BS, Stensmyr MC. 2015. Olfactory proxy detection of dietary antioxidants in *Drosophila*. *Curr Biol*. 25:455–466.
- Egea-Weiss A, Renner A, Kleineidam CJ, Szyszka P. 2018. High precision of spike timing across olfactory receptor neurons allows rapid odor coding in *Drosophila*. *iScience*. 4:76–83.
- Elsik CG, Tayal A, Diesh CM, Unni DR, Emery ML, Nguyen HN, Hagen DE. 2016. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Res*. 44:D793–D800.
- Gabler S, Soelter J, Hussain T, Sachse S, Schmucker M. 2013. Physicochemical vs. vibrational descriptors for prediction of odor receptor responses. *Mol Inform*. 32:855–865.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol*. 17:1648–1657.
- Goldman AL, Van der Goes van Naters W, Lessing D, Warr CG, Carlson JR. 2005. Coexpression of two functional odor receptors in one neuron. *Neuron*. 45:661–666.
- Grillet M, Campagner D, Petersen R, McCrohan C, Cobb M. 2016. The peripheral olfactory code in *Drosophila* larvae contains temporal information and is robust over multiple timescales. *Proc R Soc B*. 283:1–8.
- Haddad R, Carmel L, Sobel N, Harel D. 2008. Predicting the receptive range of olfactory receptors. *PLoS Comput Biol*. 4:e18.
- Hallem EA, Carlson JR. 2006. Coding of odors by a receptor repertoire. *Cell*. 125:143–160.
- Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA*. 89:10915–10919.
- 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.
- Hu XS, Matsunami H. 2018. High-throughput odorant receptor deorphanization via phospho-S6 ribosomal protein immunoprecipitation and mRNA profiling. In: Simoes de Souza FM, Antunes G, editors. *Olfactory receptors: methods and protocols*. New York (NY): Springer New York. p. 95–112.
- Hughes DT, Wang G, Zwiebel LJ, Luetje CW. 2014. A determinant of odorant specificity is located at the extracellular loop 2-transmembrane domain 4 interface of an *Anopheles gambiae* odorant receptor subunit. *Chem Senses*. 39:761–769.
- Jiang Y, Gong NN, Hu XS, Ni MJ, Pasi R, Matsunami H. 2015. Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat Neurosci*. 18:1446–1454.
- Jones EM, Jajoo R, Cancelli D, Lubock NB, Wang J, Satyadi M, Cheung R, de March C, Bloom JS, Matsunami H, et al. 2019. A scalable, multiplexed assay for decoding GPCR-ligand interactions with RNA sequencing. *Cell Syst*. 8:254–260.e6.
- Karpe SD, Dhingra S, Brockmann A, Sowdhamini R. 2017. Computational genome-wide survey of odorant receptors from two solitary bees *Dufourea novaeangliae* (Hymenoptera: Halictidae) and *Habropoda laboriosa* (Hymenoptera: Apidae). *Sci Rep*. 7:10823.
- Katoh K, Frith MC. 2012. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*. 28:3144–3146.
- Kepchia D, Xu P, Terryn R, Castro A, Schürer SC, Leal WS, Luetje CW. 2019. Use of machine learning to identify novel, behaviorally active antagonists of the insect odorant receptor co-receptor (Orco) subunit. *Sci Rep*. 9:4055.
- Kreher SA, Kwon JY, Carlson JR. 2005. The molecular basis of odor coding in the *Drosophila* larva. *Neuron*. 46:445–456.
- Kreher SA, Mathew D, Kim J, Carlson JR. 2008. Translation of sensory input into behavioral output via an olfactory system. *Neuron*. 59:110–124.

- 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.
- Leal WS, Choo YM, Xu P, da Silva CS, Ueira-Vieira C. 2013. Differential expression of olfactory genes in the southern house mosquito and insights into unique odorant receptor gene isoforms. *Proc Natl Acad Sci USA*. 110:18704–18709.
- Li J, Haddad R, Santos V, Bavan S, Luetje CW. 2015. Receptive range analysis of a mouse odorant receptor subfamily. *J Neurochem*. 134:47–55.
- Li H, You Y, Zhang L. 2018. Single sensillum recordings for locust palp sensilla basiconica. *J Vis Exp*. 136: e57863.
- Lin CC, Potter CJ. 2015. Re-classification of *Drosophila melanogaster* trichoid and intermediate sensilla using fluorescence-guided single sensillum recording. *PLoS One*. 10:e0139675.
- Lombardo F, Salvemini M, Fiorillo C, Nolan T, Zwiebel LJ, Ribeiro JM, Arcà B. 2017. Deciphering the olfactory repertoire of the tiger mosquito *Aedes albopictus*. *BMC Genomics*. 18:770.
- Lu T, Qiu YT, Wang G, Kwon JY, Rutzler M, Kwon HW, Pitts RJ, van Loon JJ, Takken W, Carlson JR, et al. 2007. Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Curr Biol*. 17:1533–1544.
- Mainland JD, Li YR, Zhou T, Liu WL, Matsunami H. 2015. Human olfactory receptor responses to odorants. *Sci Data*. 2:150002.
- Malnic B, Hirono J, Sato T, Buck LB. 1999. Combinatorial receptor codes for odors. *Cell*. 96:713–723.
- Man O, Willhite DC, Crasto CJ, Shepherd GM, Gilad Y. 2007. A framework for exploring functional variability in olfactory receptor genes. *PLoS One*. 2:e682.
- Marshall B, Warr CG, de Bruyne M. 2010. Detection of volatile indicators of illicit substances by the olfactory receptors of *Drosophila melanogaster*. *Chem Senses*. 35:613–625.
- Montague SA, Mathew D, Carlson JR. 2011. Similar odorants elicit different behavioral and physiological responses, some supersustained. *J Neurosci*. 31:7891–7899.
- Morel P. 2018. Grammar : grammar of graphics plotting in Matlab. *J Open Source Softw*. 3:1–4.
- Münch D, Galizia CG. 2016. DoOR 2. 0—comprehensive mapping of *Drosophila melanogaster* odorant responses. *Sci Rep*. 6:1–9.
- Nagarathnam B, Karpe SD, Harini K, Sankar K, Iftexhar M, Rajesh D, Giji S, Archunan G, Balakrishnan V, Gromiha MM, et al. 2014. DOR—a Database of Olfactory Receptors—Integrated Repository for Sequence and Secondary Structural Information of Olfactory Receptors in Selected Eukaryotic Genomes. *Bioinform Biol Insights*. 8:147–158.
- Nakagawa T, Pellegrino M, Sato K, Vosshall LB, Touhara K. 2012. Amino acid residues contributing to function of the heteromeric insect olfactory receptor complex. *PLoS One*. 7:1–9.
- Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arcà B, Arensburger P, Artemov G, et al. 2015. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*. 347:1258522.
- Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 48:443–453.
- Nichols AS, Luetje CW. 2010. Transmembrane segment 3 of *Drosophila melanogaster* odorant receptor subunit 85b contributes to ligand-receptor interactions. *J Biol Chem*. 285:11854–11862.
- Nishizumi H, Sakano H. 2015. Decoding and deorphanizing an olfactory map. *Nat Neurosci*. 18:1432–1433.
- Oka Y, Katada S, Omura M, Suwa M, Yoshihara Y, Touhara K. 2006. Odorant receptor map in the mouse olfactory bulb: in vivo sensitivity and specificity of receptor-defined glomeruli. *Neuron*. 52:857–869.
- Olsen SR, Bhandawat V, Wilson RI. 2010. Divisive normalization in olfactory population codes. *Neuron*. 66:287–299.
- Olsson SB, Hansson BS. 2013. Electroantennogram and single sensillum recording in insect antennae. In: Kazushige T, editor. *Pheromone signaling: methods and protocols*. Totowa, NJ: Humana Press. p. 157–177.
- Omasits U, Ahrens CH, Müller S, Wollscheid B. 2014. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*. 30:884–886.
- Raman B, Joseph J, Tang J, Stopfer M. 2010. Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J Neurosci*. 30:1994–2006.
- Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner DS, Pisani D, Dekker T, Anfora G, Rota-Stabelli O. 2016. The evolution of olfactory gene families in *Drosophila* and the genomic basis of chemical-ecological adaptation in *Drosophila suzukii*. *Genome Biol Evol*. 8:2297–2311.
- Ray A, van Naters WG, Carlson JR. 2014. Molecular determinants of odorant receptor function in insects. *J Biosci*. 39:555–563.
- Ronderos DS, Lin CC, Potter CJ, Smith DP. 2014. Farnesol-detecting olfactory neurons in *Drosophila*. *J Neurosci*. 34:3959–3968.
- Saberi M, Seyed-Allaei H. 2016. Odorant receptors of *Drosophila* are sensitive to the molecular volume of odorants. *Sci Rep*. 6:1–11.
- Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. 2009. Odor coding by a mammalian receptor repertoire. *Sci Signal*. 2:ra9.
- Schmucker M, de Bruyne M, Hähnel M, Schneider G. 2007. Predicting olfactory receptor neuron responses from odorant structure. *Chem Cent J*. 1:11.
- Spors H, Wachowiak M, Cohen LB, Friedrich RW. 2006. Temporal dynamics and latency patterns of receptor neuron input to the olfactory bulb. *J Neurosci*. 26:1247–1259.
- Su CY, Martelli C, Emonet T, Carlson JR. 2011. Temporal coding of odor mixtures in an olfactory receptor neuron. *Proc Natl Acad Sci USA*. 108:5075–5080.
- Touhara K, Sengoku S, Inaki K, Tsuboi A, Hirono J, Sato T, Sakano H, Haga T. 1999. Functional identification and reconstitution of an odorant receptor in single olfactory neurons. *Proc Natl Acad Sci USA*. 96:4040–4045.
- Viklund H, Elofsson A. 2008. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*. 24:1662–1668.
- Wang G, Carey AF, Carlson JR, Zwiebel LJ. 2010. Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc Natl Acad Sci USA*. 107:4418–4423.
- Wicher D, Schäfer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS. 2008. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature*. 452:1007–1011.
- Wright GA, Thomson MG, Smith BH. 2005. Odour concentration affects odour identity in honeybees. *Proc Biol Sci*. 272:2417–2422.
- Xu P, Leal WS. 2013. Probing insect odorant receptors with their cognate ligands: insights into structural features. *Biochem Biophys Res Commun*. 435:477–482.
- Yarali A, Ehser S, Hapil FZ, Huang J, Gerber B. 2009. Odour intensity learning in fruit flies. *Proc Biol Sci*. 276:3413–3420.