Olfactory receptors of Drosophila are sensitive to molecular volume of odorants

Majid Saberi, **Hamed Seyed-allaei** (hamed@ipm.ir)

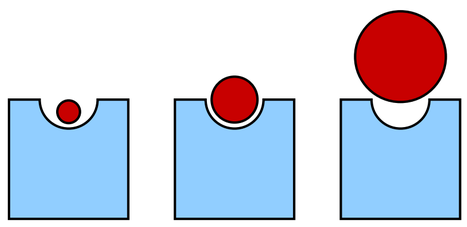
Institute for Research in Fundamental Sciences

Abstract

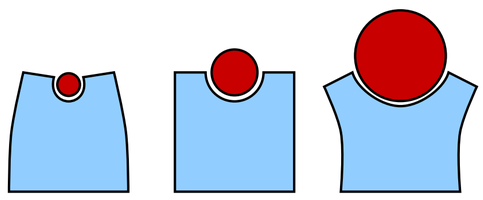
Which properties of a molecule define its odor? This is a basic question of olfaction, yet to be answered. Human olfactory system has a repertoire of about 350 olfactory receptors. Molecules bind to them with different affinities and activate them with different efficacies, resulting in a combinatorial code that identifies odorants. We hypothesized that the binding affinity between a pair of odorant-receptor is affected by their relative sizes. The affinity can reaches its maximum if molecular volume of an odorant matches volume of a receptor's binding-pocket and it reach zero if the sizes are too different, obscuring the effect of other molecular properties. We formulated this hypothesis mathematically and verified it on Data of Drosophila, and predicted the volume and the structural flexibility of each receptor's binding-site, which are significantly different among receptors. This provides a reason for differences in smell among similar molecules of different sizes.

Assumptions

Different scenarios that may happen when an odorant molecule meets a receptor:



From left to right, misfit because of small volume of molecule, perfect match and misfit because of large molecular volume.



The flexibility of a receptor may compensate for the volume mismatches. The red disks are odorant molecule, and the blue shapes are olfactory receptor and binding-pocket.

Let assume the response r_{nm} depends on the molecular volume of the odorant, v_m, and other physio-chemical properties of the molecule m;

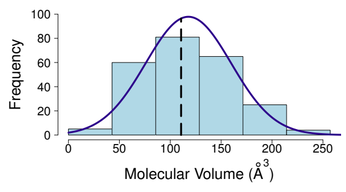
We assume that we can separate the response r_{nm} into two terms:

$$r_{nm} = f_n(v_m)\psi_{nm}. \quad (1)$$

The first term, $f_n(v_m)$, depends only on the molecular volume of odorants. The second term, ψ_{nm} include every other influential properties of molecules, but the molecular volume. Both terms are characteristic of each receptor, and they might vary from neuron to neuron.

Materials

We take the neural data of DoOR database [2] and we calculate molecular volume using a computational chemistry software – VEGA ZZ [3]. We used GNU R to analyze the data. DoOR database can be summarized in an $N \times M(55 \times 251)$ matrix. Its elements, r_{nm}, are the response of neuron n to odorant m. This matrix is normalized between 0 and 1 so we have $0 \leq r_{nm} \leq 1$, where 1 is the strongest response. The only problem is that this matrix has many *Not Available* (NA) values, 71%, and different neurons are excited by different set of odorants.



Density function of molecular volumes $g(v)$, considering all molecules of DoOR database. The solid line is a Gaussian fit (Eq. 5) and the dashed line shows the median, which is slightly different from the mean.

Contact Information

Address School of Cognitive Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Email hamed@ipm.ir

Methods

In fact, the first term, $f_n(v)$, is the tuning curve of neuron n in respect to the molecular volumes, it can be approximated with a Gaussian function

$$f_n(v) = e^{-\frac{(v-v_n)^2}{2\sigma_n^2}}, \quad (2)$$

where, v_n is the preferred molecular volume of receptor n and σ_n represents its flexibility. Here we want to estimate v_n and σ_n. To do so, first we calculate the response weighted average of molecular volumes, $\frac{\sum_m v_m r_{nm}}{\sum_m r_{nm}}$ and then we use (1):

$$\frac{\sum_m v_m r_{nm}}{\sum_m r_{nm}} = \frac{\sum_m v_m f_n(v_m)\psi_{nm}}{\sum_m f_n(v_m)\psi_{nm}}. \quad (3)$$

Here we can approximate $\sum$ with $\int$, which is common in statistical physics:

$$\sum_m \ldots f_n(v_m)\psi_{nm} \approx \langle\psi_{nm}\rangle_m \int_0^\infty \ldots f_n(v)g(v)dv. \quad (4)$$

In which, $\langle\psi_{nm}\rangle_m$ denotes the average of ψ_{nm} over all $m : r_{nm} \neq$ NA. It can be moved out of the integral for it is independent of v. In the above equation, $g(v)$ is the density of states, $g(v)dv$ indicates how many molecules have a molecular volume in the range of v and $v + dv$. This function can be approximated by a Gaussian function:

$$g(v) = e^{-\frac{(v-v_g)^2}{2\sigma_g^2}}, \quad (5)$$

ideally, $g(v)$ should not depend on the neuron n, it is the property of ensemble of odorant molecules, not neurons. But here, we have many missing values ($r_{nm} = NA$), so we have to calculate $g(v)$ for each neuron separately; Therefore, v_{g_n} and σ_{g_n} are the average and standard deviation of molecular volume while $r_{nm} \neq$ NA. Now we rewrite equation (3) using equation (4):

$$\frac{\sum_m v_m r_{nm}}{\sum_m r_{nm}} \approx \frac{\int v f_n(v) g_n(v) dv}{\int f_n(v) g_n(v) dv}. \quad (6)$$

We replace the product of $f_n(v)$ and $g_n(v)$ in the above equation with $h_n(v) = f_n(v)g_n(v)$, to make a simpler form

$$\frac{\sum_m v_m r_{nm}}{\sum_m r_{nm}} \approx \frac{\int_v v h_n(v) dv}{\int_v h_n(v) dv}. \quad (7)$$

The function $h_n(v)$ is a Gaussian function because it is the product of two Gaussian functions,

$$h_n(v) = e^{-\frac{(v-\mu_{h_n})^2}{2\sigma_{h_n}^2}}, \quad (8)$$

so the right hand side of equation 7 is nothing but μ_{h_n} and in a similar way, we can calculate σ_{h_n} from the neural data

$$\mu_{h_n} \approx \frac{\sum_m v_m r_{nm}}{\sum_m r_{nm}}, \sigma_{h_n}^2 \approx \frac{\sum_m v_m^2 r_{nm}}{\sum_m r_{nm}} - \mu_{h_n}^2 \quad (9)$$

We knew the mean v_{g_n} and standard deviation σ_{g_n} of $g_n(v)$ from the molecular volumes of the ensembles of odorants. We just calculated the mean μ_{h_n} and standard deviation σ_{h_n} of $h_n(v)$ from the neural data. Now calculating the mean v_n and the standard deviation σ_n of $f_n(v)$ is trivial, first we calculate σ_n from

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_{h_n}^2} - \frac{1}{\sigma_{g_n}^2} \quad (10)$$

and then we calculate v_n:

$$\frac{v_n}{\sigma_n^2} = \frac{\mu_{h_n}}{\sigma_{h_n}^2} - \frac{v_{g_n}}{\sigma_{g_n}^2}. \quad (11)$$

Tests

We calculate p-values by permutation test.

$$Alternative : \sigma_n \neq \infty \quad (12)$$

$$Null : \sigma_n \to \infty \quad (13)$$

so p-value is the probability of having $\sigma'_n \leq \sigma_n$, where σ_n is calculated from the original data, but σ'_n calculated using permuted version.
We are testing a hypothesis on ∼60 olfactory receptors simultaneously. If we use a simple threshold of 0.05 for the p-value of each receptor, we may have false positives. To address this issue, multiple-comparison problem, we use Bonferroni correction. The problem with Bonferroni correction is that it may increases false negatives. This problem can be addressed using another method – False Discovery Rate (FDR). We used both methods – Bonferroni and FDR, as well as no correction.
We also want to show the diversity of volume and flexibility of binding pocket among receptors. To estimate the p-values, we take any pair of receptors that was sensitive to molecular volume (28 receptors), calculate their difference, use a permutation test and measure the probability of being different only by chance.

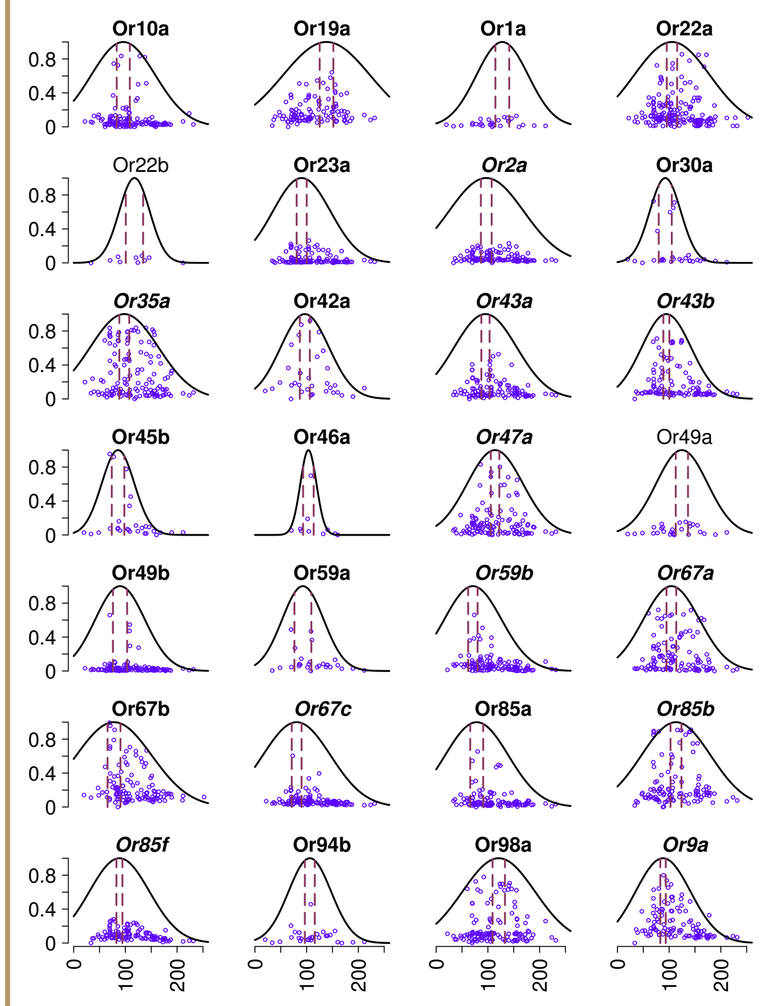
References

[1] Majid Saberi, Hamed Seyed-allaei, *Olfactory receptors are sensitive to molecular volume of odorants.*, 2015, bioRxiv, DOI:http://dx.doi.org/10.1101/013516

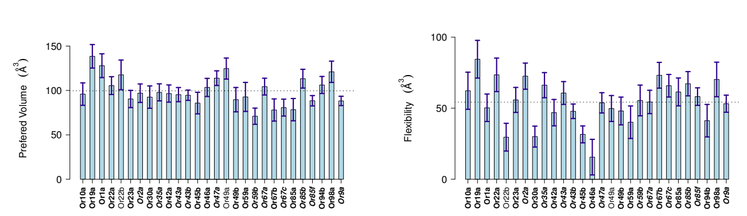
[2] Galizia CG *et al.*, *Integrating heterogeneous odor response data into a common response model: A DoOR to the complete olfactome.*, 2010, Chemical Senses, DOI:10.1093/chemse/bjq042

[3] Pedretti *et al.*, *VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming.*, 2004, Journal of Computer-Aided Molecular Design, DOI:10.1023/B:JCAM.0000035186.90683.f2

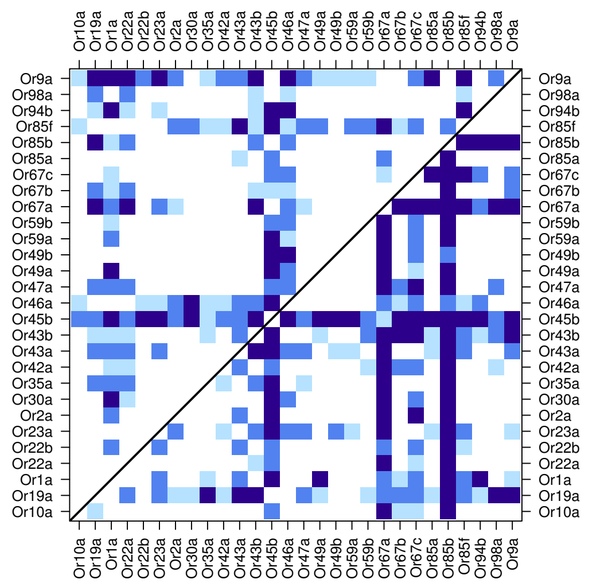
Results



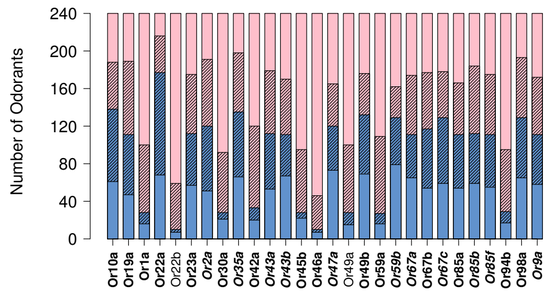
Response of olfactory receptors versus molecular volume of odorants (circles), the fitted functions $f_n(v)$ from Eq. 1 (solid lines), and the error bars of the mean of $f_n(v)$ (red vertical lines), for 28 receptors that their response showed significant (p-value < 0.05) dependence to molecular volume. Among them, 26 are significant according to FDR correction (receptor names in bold) and 11 are significant considering Bonferroni correction (receptor names in italic).



The preferred volumes of 28 receptors v_n (left). and their flexibilities σ_n (right). The error bars are calculated using Jack-Knife method.



Pairs of olfactory receptors that differ significantly in their binding-pocket's volume (upper triangle) and flexibility (lower triangle). All blue shades have p-value of less than 0.05, two darker shades have FDR corrected p-values of less than 0.05 and the darkest shade has Bonferroni corrected p-value of less than 0.05.



Venn diagram of DoOR database and our suggested important odorants of each receptor. The database includes 240 molecules, some are used to study an olfactory receptor (blue areas), and data for the rests are not available (pink). The hatched area are odorants with molecular volume close to the preferred volume of each receptor ($v_n \pm \frac{\sigma}{2}$). We already know the neural response of hatched blue areas, but the hatched pink odorants could be the target of future experiments, we predict the rest give only zeros.