Novel paradigm for odorant-olfactory receptor interactions

Abstract
This opinion paper recommends a novel methodology to mechanistically elucidate the interactions between olfactory receptors and odorant molecules. Preliminary results presented here show diffusion of odorant-ligand molecules from outside the olfactory receptor protein into the receptor binding region. This notion marks a sharp contrast with earlier computational studies related to olfactory receptor-odorant interactions. These have drawn conclusions from static or dynamic studies that have started from the ligand already docked in a preferred configuration in the receptor’s binding region. Dynamic simulations studies that begin by tracking the approach of the odorant from outside the protein, its transit to the receptor binding region, followed by tracking the rest of the receptor-odorant interactions will allow the determination of protein structural changes: 1) that facilitate the odorant path; 2) that facilitate specific interactions between amino acids and the odorant; 3) that identify the activation state of the protein and its transition to the inactive state. The novel methodology will not only improve comparisons of hypotheses generated from these computational studies with experimental studies, it will also provide novel mechanistic insights into these interactions. The primary goal is to establish a predictive paradigm for olfactory receptor-odorant interactions.

Keywords: olfactory receptors, odorants, protein modeling, protein-ligand docking, GPCRs, rhodopsin, beta-2-adrenergic receptor

Abbreviations: GPCR, g (TP-binding) protein coupled receptors; HMM, hidden markov models; OR, olfactory receptor(s); NMR, nuclear magnetic resonance; TM, transmembrane; RMSD, root mean square displacement; RMSF, root mean square fluctuations

Opinion
The publication of the discovery of olfactory receptors (OR) in 1991 sparked burgeoning research efforts in the domain of chemoreception. The need to study the role of olfactory receptors in olfaction became imperative, following the publication of the initial drafts of the human genome. Olfactory or olfactory-like receptors mined from the human genome constitute its largest family, possessing between 350 and 400 functional OR genes. 2–4

Interestingly, the functional genes in the OR super family were a little more than third of the approximately, 1000 olfactory receptor repertoire mined from the genome that remain non-functional or pseudogenes. This raises very important questions as to how the sense of smell in human beings has evolved. Rodent5 and canine OR repertoires (consisting of about 1200 OR genes), when mined from their genomes, showed that approximately, half of these genes were pseudogenic, or otherwise, non-functional. More recently, the OR repertoire from the African elephant genome listed 5000 ORs, of which, 40% are functional. 6

One of the primary challenges in the study of ORs is elucidating the mechanism(s) by which a few hundred ORs in mammals discriminate several thousand odorants from the environment, which singly or in combination produce distinct and identifiable odors. OR-odorant interactions are promiscuous: one OR is known to bind multiple odorants; odorants, in turn, are known to interact with more than one OR.

Research in deorphanizing ORs-identifying odorants or odors that are likely to bind an OR-has adopted a two-pronged approach. The primary approach is experimental, where excitatory responses are identified following the exposure of olfactory receptor neurons (one receptor corresponds to one neuron) to a panel of odorants. 9–12 Excitatory responses that arise from these interactions over a range of concentrations of odorants (typically organic chemical compounds) are then measured.

This opinion paper is primarily concerned with the secondary-computational-approach. This approach has often been a companion to the experimental functional analysis approach. 13,14 But it is increasingly coming into its own as an independent sub-domain of study. Computational studies involve studying the interactions between ORs and odorants, in silico. This paper will culminate in the recommendation to revisit the established paradigms related to the computational approaches for deorphanizing ORs.

The challenges facing in silico approaches
ORs are membrane bound proteins. They belong to a class of fairly ubiquitous proteins called G(TP-binding) Protein Coupled Receptors (GPCRs). These receptors, following excitatory responses to stimuli (an interaction with an odorant or odor) bind to a G-protein and initiate a cytoplasmic signal transduction cascade catalyzing a sensory response (identification of the odor). ORs belong to Class A GPCRs, which, structurally consist of seven transmembrane (TM) helical regions, with intra- and extra-cellular loops linking these helices. Typically, GPCRs have an extracellular N-terminus and a cytoplasmic C-terminus. The C-terminal coil contains certain polypeptide amino acid motifs that are believed to be responsible for G-protein coupling. Instances for functional ORs have been found that deviate from the typical Class A GPCR structures. 15
Experimental functional analyses of ORs are beset by the typical problems related to membrane proteins. These include issues with protein-expression in traditional expression systems and also separating the membrane from the protein during purification. The latter is one problem, which along with solubilizing the protein, has, thus far, prevented an OR from being crystallized as a precursor to determining its structure. Similar problems stymie the crystallization of most membrane-bound proteins.

**OR protein model building**

The lack of a crystal or solution (NMR)-based structure of an OR means that protein models of olfactory receptors have to be created using a semi-empirical or ab initio (or a combination of the two) methods to determine its structure. The first step is to create an OR protein model by using homology modeling approaches, even if as a first step. The template protein structures that have been mostly used are the crystal structures of rhodopsin\(^{16–20}\) and the beta-2-adrenergic receptors.\(^{21,22}\) Approximately, fourteen GPCRs a potential template for a starting OR model have been crystallographically, structurally studied—many of these purified, solubilized and crystallized by co-crystallization with fairly large ionic compounds.

Secondary structure prediction steps have to be carried out to identify regions of the primary sequence that are likely to be found in loops and which are likely to form the helical regions that will introduce the cell-membrane. Hidden Markov models (HMMs) have been used to identify TM helical regions. Several on-line resources are available which use HMMs to identify secondary structures in membrane protein.\(^{23,24}\) The user inputs the protein sequences. The output identifies regions within the sequence that are likely to be helices, intracellular and extracellular loops.

A critical challenge in relying solely on homology modeling methods to determine OR models using structurally resolved GPCRs is that GPCRs have poor sequence similarities and poorer sequence identities with ORs. Sequence similarities of both Rhodopsin and the beta-2-adrenergic receptors with most ORs studied is about 30%. It is not sufficient, therefore, to use the GPCR template as the sole determinant for creating an OR model. Certain biochemical imperatives for the positioning of the helices also have to be met. For function-odorant interactions to occur, the interior of the GPCR helical domain has to be hydrophilic, while the hydrophobic side of a typically amphipathic helix has to be on the outside, in contact with the surrounding hydrophobic regions of the plasma membrane.

As a result, homology modeling is used, merely as a first step. This is to position the helices so that their geometries match those of Class A GPCRs. In OR modeling, a high resolution template structure is not any more advantageous that using one of modest resolution. High resolution structures have greater precision in positioning side chains, but do not significantly affect the positioning of the helices. One highly popular protocol for OR model building developed at CalTech University, Memstrukt,\(^{25}\) used a 6Å resolved electron diffraction structure or rhodopsin to position OR helices. This protocol persisted with this structure even when a 2.2Å\(^{26}\) resolved structure of rhodopsin became available.

Poor sequence similarities contribute certain sequence-specific structural artifacts in template GPCR structures. If these artifacts persist in the resulting OR model, they are likely to influence the results of in silico odorant binding. For example, the high resolution structure of rhodopsin has a bend in the cytoplasmic side of the seventh TM helix that is nearly orthogonal to the rest of the TM helices. There is no evidence for such a bend in the OR. OR helix lengths are far shorter than those for rhodopsin. To remove these template-specific structural artifacts, canonical (generalized) helices have been used while adhering to the TM helical geometries in GPCRs. Alternatively, each OR TM helix that results from homology modeling is “relaxed” from its template-structure artifacts using molecular dynamics simulation in a hydrated environment before being reinserted into the OR helical scaffold.\(^{27,28}\)

Once the helical model is built, the helices have to be rotated such that the hydrophilic side of the helix is pointed inwards towards the putative binding region, while the hydrophobic side is pointed outwards. Different approaches have been used to rationalize these helical rotations. One approach has been to determine effective hydrophobicities around the helix (by treating a helix as a helical wheel). These effective hydrophobicities are calculated from amino acid hydrophobic moments (different methods have been used to determine these).\(^{29,30}\) The side (or angle on the helical wheel) where the highest hydrophilicity residues is positioned such that it points into the interior of the helical scaffold. Another approach is one that uses energies. All the helices are rotated about a vertical axis passing through the length of the helix through set angle interval. At each rotation, the system energy is determined.\(^{26,31–35}\) The combination of helical rotations for which the system energy is minimum is identified as the best possible helical arrangement.

GPCR template loops are shorter than OR loop regions, while their helices are longer than OR helices. Loop positioning from homology modeling, and especially, after helix rotations, results in unnatural strains.\(^{27}\) During the helical scaffold building stage and the helix-rotation stages of model-building, the loop regions are removed from the protein sequences. These are then re-attached using loop-building strategies, identifying cysteine disulfide bridges if they exist in proximate loops.

**Odorant docking**

Once a protein OR model is in place, a computationally derived model of an odorant is docked into the OR putative binding region. While most studies have shown a preferred binding region for ORs is bound by TM helices 3, 4, 5 and 6,\(^{27,31,33–35}\) another less preferable binding region has also been identified, bound by helices 1, 2, 6 and 7.\(^{27}\) A computational study showed that docking in this region is likely to be inhibitory, while binding in the preferred region is likely to lead to OR excitation.

**OR-odorant binding**

Initial computational studies made conjectures as to odorant binding from static docking. Following docking, the proximity of the odorant in the OR binding region was observed. Amino acid residues within specified (say, hydrogen-bonding) distances were determined as crucial to binding—and any subsequent OR activation.\(^{25,31–33,35}\) Later, studies of dynamic OR-odorant interactions were published.\(^{25,27,28,29,34}\) These studies were driven largely by the availability of molecular dynamics simulation software as well high performance computing facilities.

Studies of dynamic interactions are advantageous to static docking and represent a major advancement in the field of computational interactions. Indeed, static docking becomes the starting point or \(t=0\) of a molecular dynamic simulations. While static docking provides a snap-shot of an interaction, dynamic simulations allow one to observe how OR-odor interactions evolve over time. More importantly, one
can identify which amino acid residues are more likely to influence odorant-binding for what length of time. Such a critical interaction will likely be missed if the (typically-used) low-energy docking configuration of the odorant within the binding pocket is identified as the determinant of OR-interaction.

Another advantage that dynamic simulation offers is that it traces the transit path of the odorant within the binding region of the OR. This path will offer some clues as to how this odorant will enter, bind with, and exit the OR (a novel method to address this is the thrust of this paper). One can also identify which amino acids facilitate this transit, which residues retard this transit and which interactions are responsible for eventual OR excitation. Over the duration of the simulation, the parameters of the root mean square deviations (RMSDs) and root mean square fluctuations (RMSFs) can be used to calculate how the structure of the OR changes as its interaction with the odorant evolves.

Depending on the granularity of the simulation, a lot of additional information can be gleaned from these dynamic studies. Studies of OR-odorant interactions, whereby every atom is parameterized in terms of its interactions with its neighbors, have shown that the positional modifications, conformational and configurationally changes of every atom, every side chain, the helices, and, indeed, the entire protein molecule can be tracked over the duration of the simulation. It has been averred that the OR excitation following odor binding can be identified through a structural change in the protein. Structures for the activated and inactivated rhodopsin and the beta-2-adrenergic receptors have been solved and are publicly available.16,19,21,22,26,36 Rhodopsin (retinal) and beta-2-adrenergic receptor (epinephrine and adrenaline) ligands are already embedded in the binding region. When function is desired, these ligand form Schiff base covalent bonds with specifically identified amino acid residues thereby activating the proteins. The critical differences between ligand-binding for these GPCRs and ORs is that ORs are myriad in number, they interact with several thousand odorants or odors, which approach the receptor from the external environment.

Drawing hypotheses from computational studies

Computational studies are crucial: they offer molecular insights into OR-odorant interactions. While experimental functional analyses are a true representation of events leading up to OR activation and subsequent olfaction, these are mechanistic black boxes. Computational methods to elucidate OR activation following odorant binding, admittedly, are fraught with challenges and uncertainties. One of these is the absence of an experimentally-derived structure of an OR. The other is mimicking these interactions in a realistic environment.

Computational studies to understand OR-odorant interactions have come a long way, evolving from: designing canonical helices without loop regions during OR protein model building,35 to adding loops to the protein model,34 to simulating a plasma membrane of several hundred molecules around the TM regions,25,37 to static docking of the ligand, to performing the first dynamic simulations in vacuo, while restraining the alpha-carbon atoms of the helices to ensure protein-model integrity, to finally performing simulations of a OR-odorant complex in a simulated lipid bilayer in a simulation box surrounded by close to a hundred thousand water molecules, where every atom is parameterized.27 The latter is the state of the art in this field of study.

Also, serious efforts towards computationally building OR protein models have resulted in the establishment of rigorous protocols that seek to ensure that the lack of an experimentally determined structure is not a hindrance to the effort. Significant efforts are made to ensure that structural artifacts resulting from template GPCR structures are removed and the lack of sequence similarity between ORs and GPCRs do not impact potential odorant binding. Despite this, possible gray areas associated with mimicking a true in vivo environment do persist.

To combat this, researchers perform computational studies by testing ORs and odorants that have been previously functionally assessed. These are odorants, at various concentrations, that bind, excite, inhibit or otherwise do not influence ORs. By performing computational binding studies, researchers hope to: 1) provide a mechanistic view into these interactions at a molecular level for which experimental data is already available; and, 2) by elucidating the mechanisms of OR-odorant interaction, extend these notions to OR-odorant complexes that have not been tested. The overarching aim is to establish a predictive paradigm for OR-odorant interactions.

Because no predictive mechanism exists, to identify a “rational” panel of odorants that is likely to interact with an OR, several thousand ORs have to be tested, narrowing the number of tests odorants until a few that show excitatory responses are identified. This is cost-prohibitive. Ideally, if one was to be able to match sequence-determined characteristics of the OR with a family of odorants representing a specific functional group, for example, it would significantly help the functional analysis effort.

Some interesting aspects of OR-odor interactions are worth mentioning here. Almost all of the studies are related to odorants that have charged functional groups, such as esters, ketones, aldehydes,11 alcohols and, to a lesser extent, carboxylic acids.9,10,12 Very little is known of possible interactions between non-polar organic compounds such as benzene or certain alkanes, which have distinct odors. Also, interestingly, some odorants are broadly tuned, showing excitatory responses to a broad spectrum of compounds over a range of functional groups. Other ORs,37 sometimes within the same OR family will interact with, at most, two odorants, of the large number of odorants tested.

Re-addressing the OR-odorant interaction paradigm

The methodological aspects of computational OR-odorant interactions are applicable to all GPCRs, and one hopes, will serve as a template to further research efforts in this field. Thus far, every study that has reported computational efforts towards understanding OR-odorant or GPCR-ligand interactions have been premised on the odorant or ligand already bound in the protein binding region. In reality however, odorants in the environment enter the nostril, access the OR extracellularly, interact with the OR, elicitng excitation and a sensory response, and exit the OR. Interaction studies that make conjectures to binding or are the starting point of a dynamic simulation with a pre-bound odorant essentially begin from a faulty premise.

Such previous studies (including those led by this papers authors) have completely ignored a critical part of the OR-odorant interaction: how an odorant enters the protein and transits to the binding region. It is critical to our understanding of these interactions how the receptor protein initiates, facilitates odorant entry and transit. Once binding has taken place, methodological notions and metrics to study these interactions have been well-researched. This opinion paper represents a recommendation that researchers revisit the OR-odorant interaction paradigm and add this important facet to their computational protocols-tracking the path of the ligand from outside the receptor region to within the receptor binding pocket. There is very little precedence for
these kinds of studies. While, one such study has been published for the beta-adrenergic receptor, as has been described previously in this paper, beta-adrenergic receptor-ligand binding and OR-odorant binding are fundamentally different.

Figures 1a and 1b illustrate preliminary efforts establishing our novel, suggested methodology. The OR that was studied was the human OR17-209. This is a well characterized OR, functionally, and from some of our previous studies, computationally. The odorant ligand in this study is the isooamylacetate (C\textsubscript{12}H\textsubscript{14}O\textsubscript{2}). Functional studies have established this odorant as strongly excitatory for the hOR17-209 receptor. The figures represent the simulation box that is used for our molecular dynamics simulations. This box consists of a model of the receptor in yellow. The gray molecules are several hundred 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) molecules, which are used to simulate the lipid bilayer, which represent the plasma membrane in which the protein is embedded. The rest of the box is composed of between 70- and 100,000 water molecules. The water molecules have been removed from the figure for clarity. At the beginning of the simulation, the isooamylacetate ligand molecules, colored green, are embedded among water molecules on the extracellular side of the receptor protein. As the all-atom parametrization simulation proceeds, we can track the positional and energy behavior each and every atom within the system. This includes tracking the path and behavior of every ligand as well as any structural change in the protein, for the protein as whole, individual TM helices and the side chains of the amino acids. Ideally, we hope to study our system, as illustrated as we successfully track a ligand (colored red) from outside the receptor to into its binding region.

Such a notion is novel; it ensures an improved mimicking of the OR functional analysis experimental conditions. Additionally, this notion will be able to answer two critical issues that previous computational studies have not been able to raise: 1) how do the extracellular parts of the receptor facilitate the protein’s path from outside the protein into its binding region—particularly, the loops? and, 2) can one simulate odorant concentration? The first has been addressed in the previous paragraph. To address the second: most functional analyses are performed using odorants over a range of concentrations; a strong correlation has been observed between concentration and level of excitatory response for ORs that do show a response. In our recommended setup, we can easily simulate a specific concentration by determining how many odorant molecules have to be randomly distributed among an estimated number of extracellular water molecules.

**Addressing potential problems in this novel approach**

The results of our preliminary studies as illustrated in the figures, have shown us specific potential problems that have to be addressed to ensure successful implementation of this new methodology: 1) simulations to ensure that one or more ligands can be tracked from outside the receptor to the receptor binding pocket means will be of significantly longer duration; 2) a full molecular dynamics simulation with all-atom parametrization will also significantly burden computational resources for the extended duration of the simulations.

In our system, the gray molecules representing the lipid bilayer plasma membrane serve two purposes: 1) to ensure that the protein TMs are embedded within the plasma membrane; and, 2) to allow every atom in the protein to be parameterized without restrictions. In previous studies where high performance computational facilities were not available, we preserved the integrity of the receptor protein by restricting the positions of the protein’s alpha carbon atoms. Later, when we carried out all-atom parametrized studies with the ligand in the binding region, the simulations were not carried out for long enough durations such that protein structural integrity was compromised. The evolving energetics of our proposed novel protocol and the longer simulation runs might result in the system overheating and consequent protein denaturation. In figure 1b, we can clearly see that the inter-helical distances in the OR have increased, when compared to the same distances in Figure 1. Clearly, as this methodology gets adopted and used, we and other researchers will propose methods to remedy this situation.

**Conclusion**

This opinion paper recommends readdressing computational methodologies to study OR-odorant interactions, to better and comprehensively understand OR odor interactions. The notions presented here are extensible to all GPCR-ligand interactions. They represent a significant improvement to the current employed methodologies. There is sufficient variability in interactions because

---

**Figure 1a & 1b** illustrate preliminary results and the feasibility of the proposed work.

The OR (in yellow) is embedded in a simulated lipid bilayer (gray). The odors (green) are distributed outside the receptor at the start of the simulation. The entire system is hydrated (water molecules are not shown for clarity). The odor molecule, colored red in both figures, is tracked through the simulation (we will track the structural and energetic behavior of every molecule in the system). The path followed by this odor-ligand is one putative path from outside the receptor into its binding region.

**Citation:** Lai PC, Crasto C. Novel paradigm for odorant-olfactory receptor interactions. MOJ Proteomics Bioinform. 2016;4(3):215–219. DOI: 10.15406/mojpbi.2016.04.00120
of the large number of OR genes and one or two orders of magnitude greater availability of organic compounds or putative odorants. Combinatorially therefore, an unlimited OR-odorant systems exists to test and fine-tune the recommended novel methods.

Acknowledgements

1. Peter C. Lai performed the preliminary modeling and simulation studies that illustrate the thrust of the work.
2. Chiquito J. Crasto developed the manuscript and led the research.

Conflict of interest

No financial or other conflicts of interest exist for the first author or the corresponding author.

References