

Research Paper


# Accelerating Ligand Discovery for Insect Odorant Receptors

Arthur Comte<sup>1,2</sup>, Maxence Lalis<sup>1</sup>, Ludvine Brajon<sup>2</sup>, Riccardo Moracci<sup>2</sup>, Nicolas Montagné<sup>2,3</sup>, Jérémie Topin<sup>1</sup>, Emmanuelle Jacquin-Joly<sup>2</sup>, Sébastien Fiorucci<sup>1</sup>

1. Université Côte d'Azur, Institut de Chimie de Nice (ICN) UMR 7272, CNRS, 06008 Nice, France.

2. INRAE, Sorbonne Université, CNRS, IRD, Université Paris Cité, Université Paris Est Créteil Val de Marne, Institut d'Ecologie et des Sciences de l'Environnement de Paris (iEES-Paris), Versailles cedex 78026, France.

3. Institut universitaire de France (IUF), France.

 Corresponding authors: sebastien.fiorucci@univ-cotedazur.fr, emmanuelle.joly@inrae.fr.© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <https://ivyspring.com/terms> for full terms and conditions.

Received: 2024.10.22; Accepted: 2025.01.25; Published: 2025.02.18

## Abstract

Odorant receptors (ORs) are main actors of the insects peripheral olfactory system, making them prime targets for pest control through olfactory disruption. Traditional methods employed in the context of chemical ecology for identifying OR ligands rely on analyzing compounds present in the insect's environment or screening molecules with structures similar to known ligands. However, these approaches can be time-consuming and constrained by the limited chemical space they explore. Recent advances in OR structural understanding, coupled with scientific breakthroughs in protein structure prediction, have facilitated the application of Structure-Based Virtual Screening (SBVS) techniques for accelerated ligand discovery. Here, we report the first successful application of SBVS to insect ORs. We developed a unique workflow that combines molecular docking predictions, *in vivo* validation and behavioral assays to identify new behaviorally active volatiles for non-pheromonal receptors. This work serves as a proof of concept, laying the groundwork for future studies and highlighting the need for improved computational approaches. Finally, we propose a simple model for predicting receptor response spectra based on the hypothesis that the binding pocket properties partially encode this information, as suggested by our results on *Spodoptera littoralis* ORs.

Keywords: Odorant Receptors, Structure-Based Virtual Screening, Behavioral Assays, *Spodoptera littoralis*

## Introduction

Various insect behaviours are intimately linked to their sense of smell at all stages of their lives. In complex and ever-changing environments, insects must efficiently filter and interpret essential cues to locate food sources, potential mates, predators or suitable egg-laying sites<sup>1</sup>. These chemical cues are detected by seven transmembrane domain (TM) odorant receptor (OR) proteins located in the membrane of olfactory sensory neurons (OSNs) within olfactory organs. Unlike the mammalian ORs, which belong to the G protein-coupled receptor (GPCR) superfamily of seven TM proteins, insect ORs are odorant-gated ion channels directly responsible for signal transduction. They assemble in tetrameric complexes composed of two subunits: a conserved

co-receptor (Orco) subunit and a highly divergent OR subunit<sup>2,3,4,5,6</sup>. The OR subunit, which houses the odorant-binding site, confers chemical selectivity to the heteromeric complex while Orco ensures subcellular OR trafficking, complex assembly, and participates in the pore architecture together with the variable OR<sup>2,5,6,7,8,9</sup>. Such an unusual ion channel structure has been confirmed by recent experimental structures obtained using cryo-electron microscopy (cryo-EM)<sup>2,3,5,6</sup>.

Over the past two decades, the sequencing of a large number of insect transcriptomes and genomes has provided a wealth of OR sequences<sup>10</sup>. The number of ORs genes varies significantly among species, ranging from as few as 4 in damselflies<sup>11</sup> to several

hundred in social insects<sup>12</sup>. These genes evolve rapidly through a dynamic process of duplication, divergence, pseudogenization, and loss. This "birth-and-death" process results in the expansion or reduction of the OR repertoire reflecting the adaptation of insects to a diverse range of ecological niches<sup>13</sup>. Although ligands have now been identified for a large list of insect ORs using various heterologous expression systems<sup>10</sup>, ligand identification of many more ORs is needed to understand how a given species mobilizes its OR repertoire to interact with its environment. This will enhance our understanding of the molecular evolution of these ecologically important insect receptors.

Ligand-Based Virtual Screening (LBVS) represents a promising way to accelerate this process, by screening large databases and identifying the physico-chemical properties of an odorant required for activating a given OR<sup>14,15,16,17,18,19</sup>. Computational models can rapidly screen extensive libraries of relevant molecules, and the identified hits can move on to *in vitro* or *in vivo* testing for further validation. Traditional insect OR functional analyses typically limit the selection of odorants tested to those found in the animal's environment<sup>20,21</sup>, potentially missing certain binders. Integrating advanced computational methods into this process not only accelerates the discovery of effective molecules but also expands the pool of potential candidates. The main limitation of LBVS is its reliance on ligand knowledge, which makes it applicable only to receptors that have already been deorphanized (i.e. ligand identified). However, the recent advances in our understanding of insect OR structures<sup>2,3,5,6</sup>, combined with the cutting-edge technology of AlphaFold<sup>22</sup>, provide an opportunity to implement Structure-Based Virtual Screenings (SBVS). Relying on docking simulations, SBVS has the potential to overcome LBVS barriers and greatly expand our ability to identify OR-odorant interactions *de novo*. SBVS is well established for mammalian GPCRs<sup>23</sup>, however its application towards insect ORs, which are unrelated to GPCRs, has yet to be demonstrated. This is primarily due to significant disparities in binding site structures. By November 2023, cryo-EM had resolved the structures of 523 GPCRs<sup>24</sup>, compared to only 5 insect ORs by 2024<sup>2,3,5,6</sup>. Currently, docking simulations on insect ORs have predominantly been employed to enhance our comprehension of the molecular mechanisms underlying olfactory signal transduction across various species by pinpointing receptor residues that may interact with known ligands, thereby providing hypotheses tested through targeted mutagenesis<sup>3,5,6,25,26,27,28,29</sup>. While few studies have

begun to suggest the benefits of docking for predicting new ligands for ORs, none of them have experimentally validated their predictions<sup>30,31</sup>. To date, such complete investigations in insects have only been conducted on odorant-binding proteins (OBPs)<sup>32,33,34</sup>, a family of soluble proteins proposed to transport odorants to the OR<sup>35</sup>.

Here, we present a comprehensive study of virtual screening for insect ORs, spanning from *in silico* predictions to *in vivo* validations and behavioral assays. We screened a large database containing 407,270 natural compounds against two deorphaned ORs from the leafworm moth *Spodoptera littoralis* (Slit) with contrasting tuning breadths: SlitOR25, which exhibits a broad recognition spectrum, and SlitOR31, specifically tuned to interact with eugenol<sup>20</sup>. Recognized as a cornerstone model in chemical ecology, *S. littoralis* has been instrumental in providing valuable insights in this field for several decades. Thus, many odorants from this species' host plants are known. Nevertheless, during our analysis we avoided making any assumptions about the chemicals responsible for activating these receptors, to increase our chance of discovering new, unexpected ligands. As a result, our SBVS approach led us to discover new OR ligands, one of which was localized in a previously unexplored area of the chemical space - an area that would remain inaccessible using LBVS methods. SBVS predictions were not only experimentally verified at the OR level, but their behavioral relevance was also investigated, pinpointing new insect attractants. We also highlight the need for meticulous evaluation of raw output generated by docking software since such data often fails to distinguish directly between binders and non-binders. Ultimately, the results of our predictive workflow, demonstrated across both broadly and narrowly tuned receptors, pave the way for future studies to cover any OR within *S. littoralis* and we anticipate applying this methodology to ORs of additional insect species.

## Materials and Methods

### Cheminformatics and preparation of chemical libraries

The experimental dataset used to assess the docking protocol comprised 51 volatiles tested on various SlitORs in de Fouchier *et al.*, 2017<sup>20</sup> (Table S1). The Structure-Based Virtual Screening was conducted using the COCONUT online database<sup>36</sup>. This open collection of natural products contains 407,270 molecules (latest updates: January 2022). We filtered the molecules according to the following criteria: retaining those with a molecular weight not exceeding

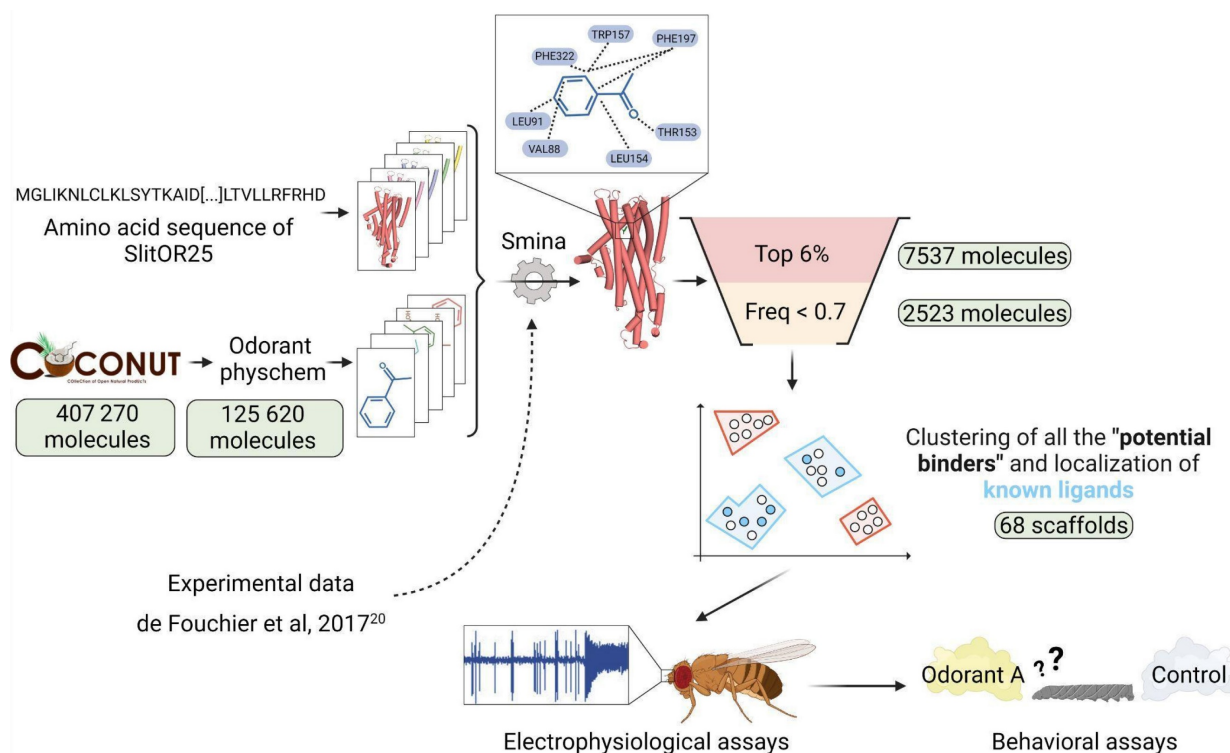
400 g/mol, featuring no more than 25 heavy atoms, having no more than 10 heteroatoms, and with a calculated octanol/water partition coefficient (logP) ranging between -1 and 7. These physico-chemical properties can be useful for distinguishing between odorous and odorless molecules, as observed in previous databases<sup>37</sup>. These molecular descriptors were calculated for all the compounds in the database using the RDKit python library (version 2023.03.3)<sup>38</sup>. Ultimately, 125,620 molecules were retained for the virtual screening (Fig. 1).

Both sets of molecules -the one from de Fouchier *et al.*<sup>20</sup> and the one from COCONUT<sup>36</sup>- were prepared using identical procedures. Gypsum-DL 1.2.0 was employed to generate 3D molecular structures from SMILES format to SDF<sup>39</sup>. All potential ionization, tautomeric, chiral, cis/trans isomeric, and ring-conformational forms were checked and optimized at a pH of  $7.0 \pm 0.5$  using Gypsum-DL<sup>39</sup>. The conversion from SDF to MOL2 format was performed using Open Babel 3.1.0<sup>40</sup>, and the structures were subsequently processed from MOL2 to PDBQT using MGLTools (version 1.57)<sup>41</sup>.

### Modeling and analysis of SlitOR structures

Five relaxed models were generated with

AlphaFold2<sup>22</sup> for the 17 receptors of *Spodoptera littoralis* studied in de Fouchier *et al.*, 2017<sup>20</sup> SlitOR3, SlitOR4, SlitOR6, SlitOR7, SlitOR13, SlitOR14, SlitOR17, SlitOR19, SlitOR24, SlitOR25, SlitOR27, SlitOR28, SlitOR29, SlitOR31, SlitOR32, SlitOR35 and SlitOR36. From the models generated for each receptor, the one with the highest pLDDT score across the entire sequence was selected for docking studies (Figure S1B). The pLDDT scoring function's reliability was validated using cryo-EM structures from previous studies<sup>2,3,5,6</sup>, confirming its suitability for evaluating model confidence (Figure S1A). Polar hydrogen atoms were added using PDB2PQR 3.6.1<sup>42</sup>, with protonation states set by PROPKA 1.0<sup>43</sup> at pH 7.0. The final structure was then minimized using the AMBER99 force field. MGLTools 1.5.7 were used to convert protein file format from PQR to PDBQT. Fpocket 4.0<sup>44</sup> was employed for cavity detection, and MDpocket<sup>45</sup> for the analysis of their physico-chemical properties. The structure of the ligand-receptor complexes was analyzed with ProLIF 2.0.3<sup>46</sup>. The reported findings were collected prior to the release of the latest AlphaFold version<sup>47</sup>. For comparative purposes, we also generated models using AlphaFold3 webserver and MODELLER 10.6<sup>48</sup>. From the five models generated with AlphaFold3 for



**Figure 1: Structure-Based Virtual Screening workflow illustrated with SlitOR25.** Starting from a large database of 407,270 natural compounds, various filtering and molecular docking steps led to a selection of less than 100 molecules for experimental validation using electrophysiological assays on the OR and behavioral assays on the larvae. Pale green labels indicate the number of molecules remaining after each step. *Odorant physchem* consists of filtering the COCONUT database<sup>36</sup> with general physicochemical properties of volatile molecules (molecular weight, number of heavy atoms, number of heteroatoms, logP)<sup>37</sup>. "Top 6%" is a selection of the top 6 percent of docked compounds that exhibit the best docking scores. "Freq < 0.7" excludes from the selection the compounds that were in the top 6 percent for more than 70 percent of the SlitORs. The clustering was performed using HDBSCAN<sup>58</sup>. Scaffolds containing known ligands are highlighted in blue, while those defining unexplored regions of the chemical space are marked in red. Refer to the methodology section for more details.

SlitOR25, SlitOR31, *Aedes aegypti* (Aaeg) OR10, *Anopheles gambiae* (Agam) OR28, *Acyrtosiphon pisum* (Apis) OR5, *Machilis hrabei* (Mhra) OR5 and *Apocrypta bakeri* (Abak) Orco, we selected the one with the highest pLDDT score (Figure S1B). Template-based models were built for AaegOR10, AgamOR28, ApisOR5, MhraOR5 and AbakOrco using all available cryo-EM structures of insect ORs<sup>2,3,5,6</sup> as templates, excluding the target itself (Protein Data Bank accession 8V02, 8V3D, 8Z9A, 7LID and 6C70). For AaegOR10, AgamOR28, ApisOR5, and MhraOR5, holo structures with bound ligands were used to enhance the accuracy of the models, whereas for AbakOrco, an apo structure was used due to the unavailability of holo structure. For each protein, five models were generated per template, and the model with the lowest DOPE assessment score was retained.

### Docking simulations

Cryo-EM structures of MhraOR5 in complex with eugenol and DEET (Protein Data Bank accession 7LID and 7LIG) were superimposed with SlitOR models using PyMOL 2.5.4<sup>49</sup>. After visual inspection of the cavities generated by Fpocket<sup>44</sup>, the primary cavity coinciding with the MhraOR5 binding site was identified as the putative binding region for each SlitOR (Figure S2C, D). Residues within 5 Å from the surface of these cavities were considered part of the binding regions. The pLDDT scores of the identified residues were evaluated to ensure a high-confidence prediction of this critical binding region (Figure S1B, C). The grid box used for docking simulations was then optimized to fit with the potential binding region of SlitORs, ensuring thorough coverage of cavity residues while minimizing their overall volume (center\_x = 10, center\_y = -1.67, center\_z = -12, size\_x = 13, size\_y = 13.9, size\_z = 15).

All the compounds were docked using the Vinardo scoring function implemented in the smina software<sup>50,51</sup> (Oct 15 2019 based on AutoDock Vina 1.1.2) with an exhaustiveness value of 20, selected for its balance between speed and high performance<sup>52</sup>. The entropic penalty is sometimes neglected by the docking scoring function or poorly estimated based on the number of rotamers in the molecule<sup>53</sup>. Consequently, molecules with greater molecular weights tend to better occupy the binding pocket, forming more contacts with the protein, leading to increased docking scores. To address this issue, one approach involves weighing the score based on molecular weight<sup>54,55</sup>, similar to how ligand efficiency is calculated (by dividing the docking score by the number of heavy atoms)<sup>56</sup>. We re-evaluated Vinardo scores by weighting them in three different ways based on the number of heavy atoms in each

molecule:

- LE score function: Vinardo Score Function / Number of Heavy atoms
- LEln score function: Vinardo Score Function / (1+ln(Number of Heavy atoms))
- LESA score function: Vinardo Score Function / (Number of Heavy atoms)<sup>2/3</sup>

ROC curves were generated for the smina outputs to evaluate the effectiveness of the score functions in distinguishing between binders and non-binders. The choice of the scoring function was determined by evaluating the docking performance, utilizing experimental data on the 17 modeled SlitORs (Fig. 1). To validate our choice, we performed an additional check by calculating the enrichment factor for all the scoring functions tested (Table S2).

### Refining docking results and classification of predictions

To propose a suitable number of novel candidate ligands for testing on SlitOR25 and SlitOR31 in *in vivo* experiments, several filters were applied to select the most relevant compounds (Fig. 1). Leveraging experimental data gathered from the evaluation of 51 molecules by de Fouchier *et al.*, 2017<sup>20</sup>, the true positive rate (TPR) to false positive rate (FPR) ratio was calculated for 15 SlitORs, namely SlitOR3, SlitOR4, SlitOR7, SlitOR14, SlitOR17, SlitOR19, SlitOR24, SlitOR25, SlitOR27, SlitOR28, SlitOR29, SlitOR31, SlitOR32, SlitOR35 and SlitOR36. We excluded from this analysis the pheromone receptors (PRs) SlitOR6 and SlitOR13, as they did not require the same score function as the non-pheromonal receptors to achieve good discrimination between binders and non-binders (Fig. 1). By varying the threshold of the top x% of molecules retained, as ranked by their docking scores according to the LE function, the optimal TPR/FPR ratio was achieved when the threshold was set at 6%, which was the median optimal threshold value across all receptors. This threshold optimizes the differential categorization between binders and non-binders (Fig. 1) and was confirmed by analyzing the enrichment factors associated with the selected scoring function (Table S2).

As a secondary filter, we excluded molecules showing remarkably strong interaction energy with almost every receptor, postulating that such promiscuity might result from a discrepancy in computational scoring rather than reflecting a biological reality. We analyzed the data from de Fouchier *et al.*, 2017<sup>20</sup> and calculated the occurrence frequency of molecules in the top 6% ranked with

their LE scores. If a molecule appeared in the list for 70% of receptors, we found that it was likely to be a non-binder. Therefore, we set the threshold at 0.7 (i.e. molecules predicted as binders for 70% of ORs) and refer to the molecules discarded during this filtration step as “suspected decoys”.

Only compounds that successfully passed both filtration steps were considered “potential binders” for the receptors tested. Throughout this study, we use the term “potential binders” to describe compounds predicted to interact with the receptor, regardless of their potential agonist or antagonist activity. Single-sensillum recordings are required to definitively distinguish between agonists and antagonists. In re-examining de Fouchier's experimental data<sup>20</sup>, we also calculated the median rank at which the TPR equals 1 for the 15 non-pheromonal SlitORs. This indicates a threshold below which all molecules are predicted to be non-binders. Consequently, we categorized molecules that fell within the lowest 45% of the docking score distribution as “potential non-binders”.

### Clustering

To thoroughly explore the extensive chemical space and maximize the likelihood of discovering novel ligands with the most diverse structures possible, “potential binders”, “potential non-binders” and “suspected decoys” identified through virtual screening of the COCONUT database<sup>36</sup> underwent separate clustering based on their molecular structures, following the same protocol. The Morgan2 fingerprints (2048-bit vector) of each molecule in the three sets were generated using the RDKit library<sup>38</sup>. Then, UMAP 0.5.4<sup>57</sup> was employed to reduce the dimensionality down to 2 with parameters set to  $n\_neighbors=10$ ,  $min\_dist=0.0$ , and default settings. This was followed by HDBSCAN 0.8.38<sup>58</sup> for data clustering with a minimum cluster size of 30 and default parameters. The consistency of intra- and inter-cluster distances was evaluated by comparing the mean Tanimoto coefficient of molecules within the same cluster to the Tanimoto coefficient between the centroids of different clusters.

### Chemicals

The availability of molecules with the best docking scores in each cluster of “potential binders”, “potential non-binders” and “suspected decoys” was verified by querying the AMBINTER supplier database. Exceptionally, for clusters of “potential binders” which contain already known ligands of SlitOR25 and SlitOR31, the availability of the top 5 molecules with the best docking scores was also checked. Selected compounds were purchased from

AMBINTER (Orléans, France) and are listed in Table S1. Depending on their solubility, odorants were either directly diluted in paraffin oil or passed through a stock solution in DMSO. We ensured that final dilution did not contain more than 5 % DMSO.

### Heterologous expression of SlitOR25 and SlitOR31 in *Drosophila melanogaster* olfactory neurons

Flies were reared on a standard cornmeal-yeast-agar medium and kept in Sanyo incubators (MIR-553) at a temperature of 25°C, under a 12-hour:12-hour light-dark photoperiod. The flies were transferred to 29°C for 24 hours before single-sensillum recording to optimize GAL4 activity while minimizing any impact on line viability<sup>59</sup>. The UAS-SlitOR25 and UAS-SlitOR31 fly lines were previously generated<sup>20</sup>. These two UAS lines were crossed with *Df(2L)Or22ab, Tl{GAL4}Or22ab* stock flies to express the OR of interest in ab3 *Drosophila melanogaster* (Dmel) sensilla instead of OR22a<sup>60</sup>.

### Single-sensillum recordings of ab3 *Drosophila* sensilla

Male and female flies, aged 2 to 5 days, were randomly selected from the *Drosophila* population. The flies were immobilized inside a 200  $\mu$ l pipette tip leaving only the head exposed. The pipette tip was affixed with dental wax onto a microscope glass slide, ensuring that the ventral side of the fly faced upwards. The antennae were secured in place using a glass capillary positioned between the second and the third antennal segment. The entire assembly was positioned under a microscope (BX51WI, Olympus, Tokyo, Japan) equipped with an SLMPLN 100X objective. A continuous flow of charcoal-filtered and humidified air at a rate of 1.5 L.min<sup>-1</sup> was directed through a glass tube with a 7 mm diameter to prevent desiccation of the flies.

The stimulation cartridges used for the screening of the odorant panel were made by placing a 1 cm<sup>2</sup> filter paper loaded with 10  $\mu$ l of each odorant solution (10  $\mu$ g/ $\mu$ l) into a Pasteur pipette. To prevent unintended evaporation of volatile compounds, a 1000  $\mu$ l pipette plastic tip was used to securely seal the Pasteur pipette. Odorant stimulations were performed by inserting the Pasteur pipette into an opening in the glass tube from which the continuous airflow was delivered, and initiating a 500 ms air pulse at a rate of 0.6 L/min.

The firing activity of ab3A OSNs was recorded using tungsten electrodes. The reference electrode was inserted into the fly's eye using a manually controlled micromanipulator (Three-Axis Coarse Mechanical Micromanipulator UM-3C, Narishige,

Tokyo, Japan). The recording electrode was inserted at the base of the recorded sensillum using a motor-controlled micromanipulator (PatchStar Micromanipulator, Scientifica, Uckfield, United Kingdom). The electrical signal was amplified using an EX-1 amplifier (Dagan, Minneapolis, Minnesota, United States of America), digitized through a Axon Digidata 1550A Data Acquisition System (Molecular Devices, Sunnyvale, California, United States of America), recorded and analyzed using the pCLAMP 10.5 software (Molecular Devices). Net responses of ab3A neurons expressing SlitOR25 or SlitOR31 were determined by subtracting the spontaneous firing rate from the firing rate during the odorant stimulation, following the methodology outlined in de Fouchier *et al.*, 2017<sup>20</sup>. To distinguish ab3 from other basiconic sensilla, 3 diagnostic stimuli were employed: 10 µg of sulcatone, 10 µg of ethyl acetate and CO<sub>2</sub> delivered by human expiration. These stimuli are potent activators of ab3B OSNs, ab2A OSNs and ab1C OSNs, respectively<sup>61</sup>. The absence of the endogenous receptor OR22a in the ab3A neuron was confirmed by employing 10 µg of ethyl hexanoate, a strong ligand of DmelOR22a. Additionally, the presence of the SlitORs in the transformed ab3A OSNs was confirmed using known ligands as positive control: acetophenone for SlitOR25 and eugenol for SlitOR31<sup>20</sup>.

For each of the two receptors, the entire panel was tested across 7 different flies. A compound was considered active if the neuronal response it elicited was statistically different from the response induced by the solvent alone (Friedman test, followed by a Dunnett's multiple comparison test adjusted with Benjamini & Hochberg correction,  $p < 0.05$ ). Dose-response experiments were conducted for all identified ligands, ranging from 1 ng to 100 µg. Data were recorded and analyzed using pCLAMP 10 and all statistical analyses were performed using R (version 4.3.2). Additionally, the predicted vapor pressure at 25°C (VP) for each tested compound was calculated using EPI Suite 4.11 to aid in the interpretation of the results.

### Larvae behavioral assays

The potential attractiveness of the new compounds that activated SlitOR25 during the electrophysiological assays were tested using Y-tube olfactometers. We did not test the new ligand of SlitOR31 because eugenol, the primary ligand for this receptor, is known to be non-attractive to the larvae of *S. littoralis*<sup>62</sup>.

Two-choice behavioral assays were performed at 24°C under dim red light using 12-15 day old larvae (third or fourth instar, L3-L4) reared on artificial diet. Caterpillars were starved the night before the

experiments (16 to 20 hours starvation). It was shown that this condition did not impact the survival or mobility of the caterpillars but increased the interest of the caterpillars in odor sources<sup>63</sup>. The olfactometer consisted of a glass Y-tube with an internal diameter of 2.1 cm. The main segment was 13 cm long and each of the two arms was 9.5 cm long. The air passing through the system was purified with charcoal. A flow meter (ProFLOW 6000, Restek, Bellefonte Pennsylvania, USA) ensured that the flow rate in each arm of the olfactometer was kept constant (0.5 L/min). Each of the tested odorants was diluted in paraffin oil at 10, 1 and 0.1 µg/µl. 10 µl of these solutions were loaded on a filter paper in one of the two arms of the Y-tube. 10 µl of paraffin oil was deposited on the filter paper in the other arm of the olfactometer. As the attractiveness of benzyl alcohol has already been demonstrated<sup>62</sup>, 100 µg of this compound was used as a positive control. Neutral controls consisted of paraffin oil in both arms, and paraffin oil versus (E)-ocimene, a compound that does not induce any *S. littoralis* larval behavior<sup>62</sup>.

Each caterpillar was tested only once, each olfactometer was changed every 4 caterpillars, and the two arms of the olfactometer were interchanged between individuals. After each experiment, all the glass parts of the device were cleaned for 1 hour in a 3-5 % solution of detergent TDF4 (Franklab, Montigny-le-Bretonneux, France), rinsed with distilled water, dried, rinsed with acetone, and put in an oven at 200 °C for 1 hour.

Pearson's Chi-squared tests were used to confirm whether the proportion of larvae preferentially choosing one arm of the olfactometer differed from 50:50 ( $p < 0.05$ ). The choice of a caterpillar was recorded if three quarters of its body was visible in one of the arms of the olfactometer within 10 minutes.

## Results

### Selecting the appropriate scoring function enhances docking performance in discriminating odorant receptor binders and non-binders

To determine the most effective scoring function for distinguishing potential binders from non-binders in the COCONUT database<sup>36</sup>, we carried out a preliminary comparison. The aim was to assess the performance of different scoring functions using experimental data from de Fouchier *et al.*, 2017<sup>20</sup>, wherein the responses of 17 *S. littoralis* ORs to 51 odorants, including plant volatiles and sex pheromone compounds, were measured. We plotted receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) for each

receptor<sup>64</sup>. First, we separated our analysis between pheromone receptors (PRs, tuned to sex pheromone compounds) and non-pheromonal receptors due to their divergent responses (Fig. 2A). Notably, the AUC values of ROC curves were markedly high for PRs when utilizing the Vinardo scoring function, with a median value of 0.90. Conversely, this value was considerably lower for non-pheromonal ORs, with a median value of 0.39. Consequently, the Vinardo scoring function could not efficiently discriminate between binders and non-binders for non-pheromonal receptors like SlitOR25 and SlitOR31. Weighting the Vinardo scoring function by considering the number of heavy atoms (LE scoring function) resulted in a remarkable improvement for SlitOR25 and SlitOR31, almost doubling the AUC values to 0.87 and 0.72, respectively (Fig. 2A-B). Such a substantial improvement was observed for all non-pheromonal receptors, with a median AUC value of 0.72. The increase in the AUC values with the LESA and LEIn scoring functions were less pronounced, reaching 0.48 and 0.65 for SlitOR25 and SlitOR31, respectively (Fig. 2A-B).

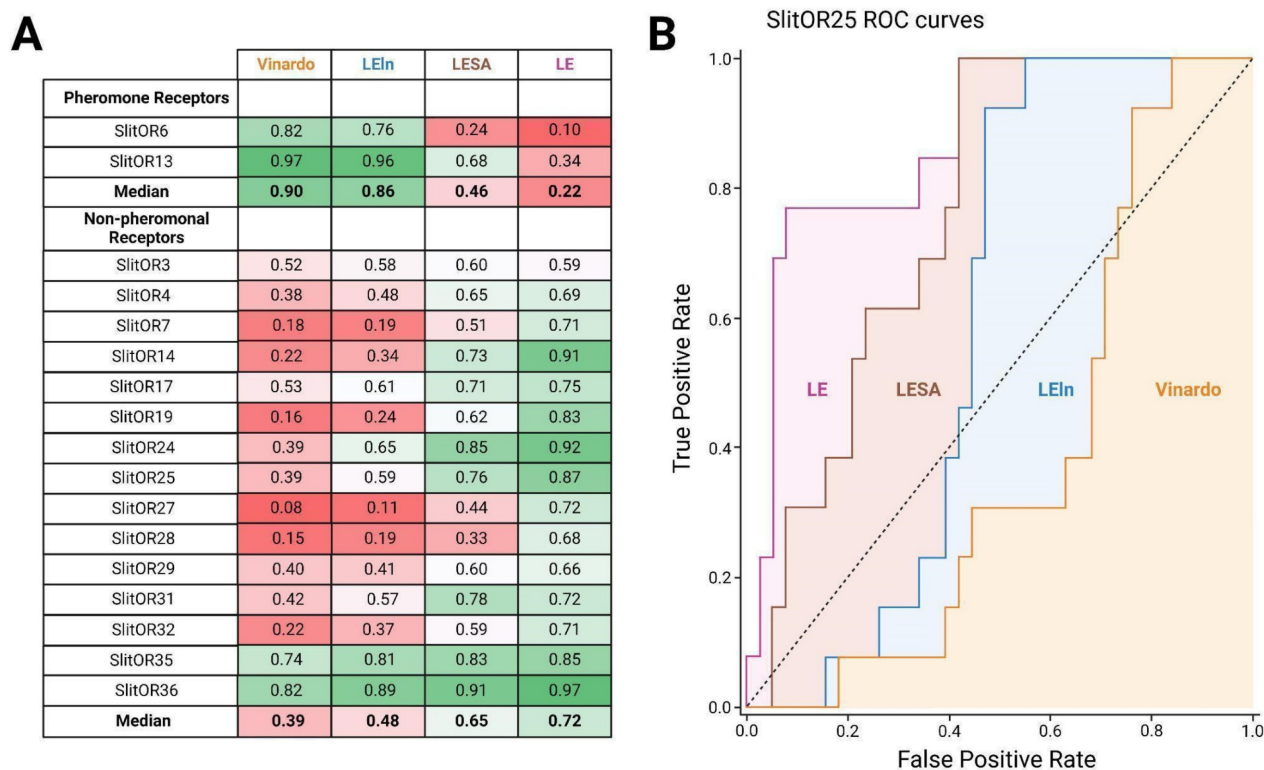
Equivalent results were obtained from the calculation of enrichment factors associated with the different scoring functions evaluated in this study.

The maximum enrichment factor values were high for PRs using the Vinardo scoring function, with a median value of 8.93, and high for other ORs using the LE scoring function, with a median value of 5.67. Notably, replacing the Vinardo scoring function with the LE scoring function resulted in a substantial improvement for SlitOR25 and SlitOR31, increasing their maximum enrichment factors by approximately threefold and sixfold, respectively (Table S2).

During the screening of the COCONUT database, the LE scoring function was then employed to assess the docking poses of the 407,270 molecules.

### Virtual screening of a large database opens unexplored chemical spaces of novel active compounds

After implementing an optimal docking procedure, we narrowed down 407,270 natural compounds to 2,523 potential binders for SlitOR25 and 2,516 for SlitOR31. These subsets consisted of the top 6% scoring compounds devoid of probable decoy molecules, i.e. removing those interacting with more than 70% of non-pheromonal receptors. The chemical space of SlitOR25 and SlitOR31 was depicted using UMAP<sup>57</sup> based on the chemical structure (Morgan2 fingerprint) of potential binders. HDBSCAN



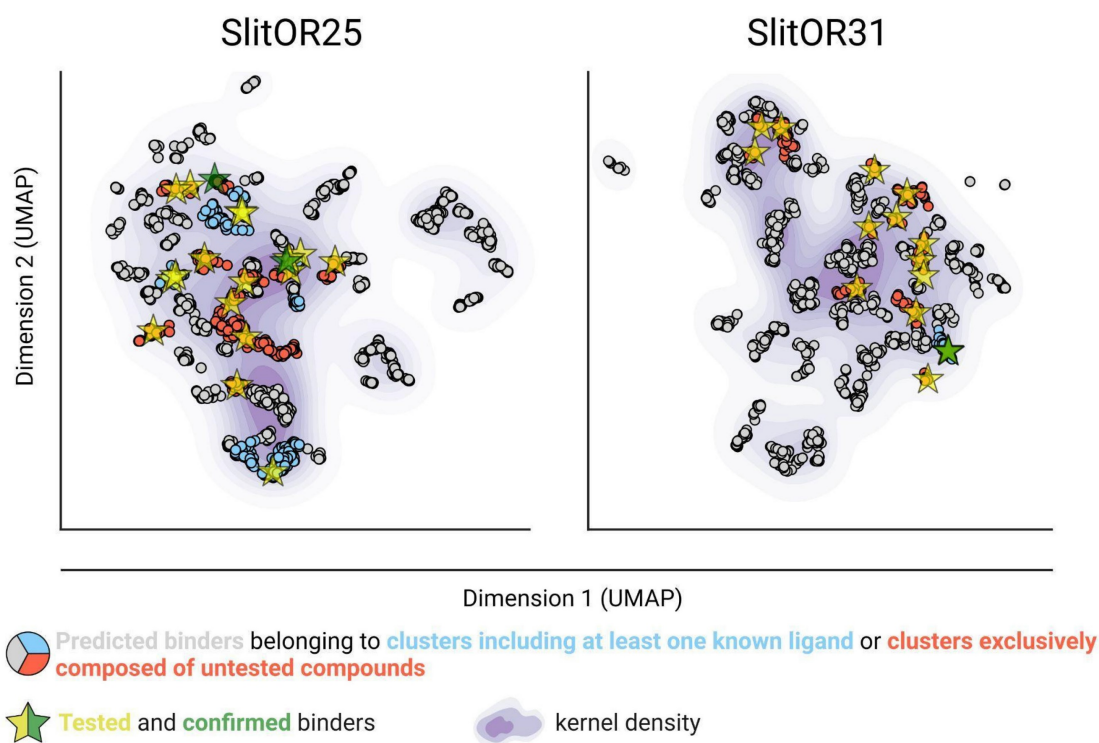
**Figure 2: LE stands out as the top-performing scoring function for analyzing the odorant docking poses on non-pheromonal odorant receptors of *S. littoralis*, including SlitOR25 and SlitOR31.** (A) The comparison table summarizes AUC values from ROC curves using Vinardo, LESA, LEIn, and LE scoring functions based on optimal poses of 51 molecules across 17 receptors from de Fouchier et al., 2017<sup>20</sup>. The LESA, LEIn, and LE functions are weighted versions of the Vinardo function as described in methodology. The gradient color scheme highlights AUC values, with green indicating higher performance and red lower ones. (B) ROC curves for SlitOR25 show performance with Vinardo (orange), LEIn (blue), LESA (brown), and LE (pink) scoring functions.

clustering<sup>58</sup> grouped these molecules into 68 clusters for SlitOR25 and 70 clusters for SlitOR31 (Figure S3). Clusters containing natural compounds with previously known experimental activity on SlitOR25 or SlitOR31 have been identified (Fig. 3). Molecules with top docking scores were chosen to represent each cluster. For experimental assays, we prioritized those purchasable and localized in the most densely populated clusters (Fig. 3). The 19 potential binders selected for testing on SlitOR25 fell into 15 clusters. Eight molecules were localized within four clusters housing known ligands, while 11 belonged to clusters never explored for SlitOR25. Among the 17 molecules picked for SlitOR31 testing, four were within the cluster housing the sole known ligand for this receptor, while the remaining ones were spread across 13 unexplored clusters. For quality-control purposes, we also selected 3 suspected decoys and 5 potential non-binders for *in vivo* testing to validate that the various filtration steps did not eliminate ligands of SlitOR25 and SlitOR31. They were chosen following a procedure similar to the one described above. Within each cluster of suspected decoys and potential non-binders, the molecule with the best docking score was retained. From this selection, decisions were based on supplier availability, with a focus on molecules within largest clusters.

### ***In vivo* experiments confirm the identification of new ligands for both the generalist SlitOR25 and the specialist SlitOR31 receptors**

To validate the *in silico* predictions, we conducted single-sensillum recordings on *Drosophila* ab3A OSNs expressing SlitOR25 or SlitOR31 in place of the endogenous DmelOR upon stimulation with the predicted binders, decoys and non-binders.

SlitOR25 exhibited significant activation in response to 2 out of the 19 molecules predicted as binders by docking (Fig. 4A): benzyl formate (labeled 16 in Fig. 4A and in Table S1) and 3-(3,4-dihydroxyphenyl)propanoic acid (10). This corresponded to a success rate (precision) of over 10%. Furthermore, with an accuracy of 0.36, the virtual screening campaign has effectively ruled out a significant number of potential non-active compounds, as none of the selected molecules predicted to be non-binders or suspected decoys activated SlitOR25. Additional metrics evaluating the performance of the workflow are summarized in Table S3. It should be noted that the median response of neurons expressing SlitOR25 to benzyl formate, at 162 spikes/s, was comparable to the response elicited by acetophenone<sup>20</sup> or by benzyl cyanide<sup>16</sup>, the best SlitOR25 ligands previously identified. Benzyl formate (16) belonged to a cluster of molecules



**Figure 3: Virtual screening allowed the identification of new ligands in unexplored areas within the predicted chemical space of SlitOR25 and SlitOR31.** UMAP representation of the chemical space based on Morgan2 fingerprints of potential binders, and with kernel density estimates illustrated using a purple gradient. All the potential binders experimentally tested are represented by yellow stars. These stars are recolored in green when the tested compound is confirmed to be a real ligand for the receptor of interest.



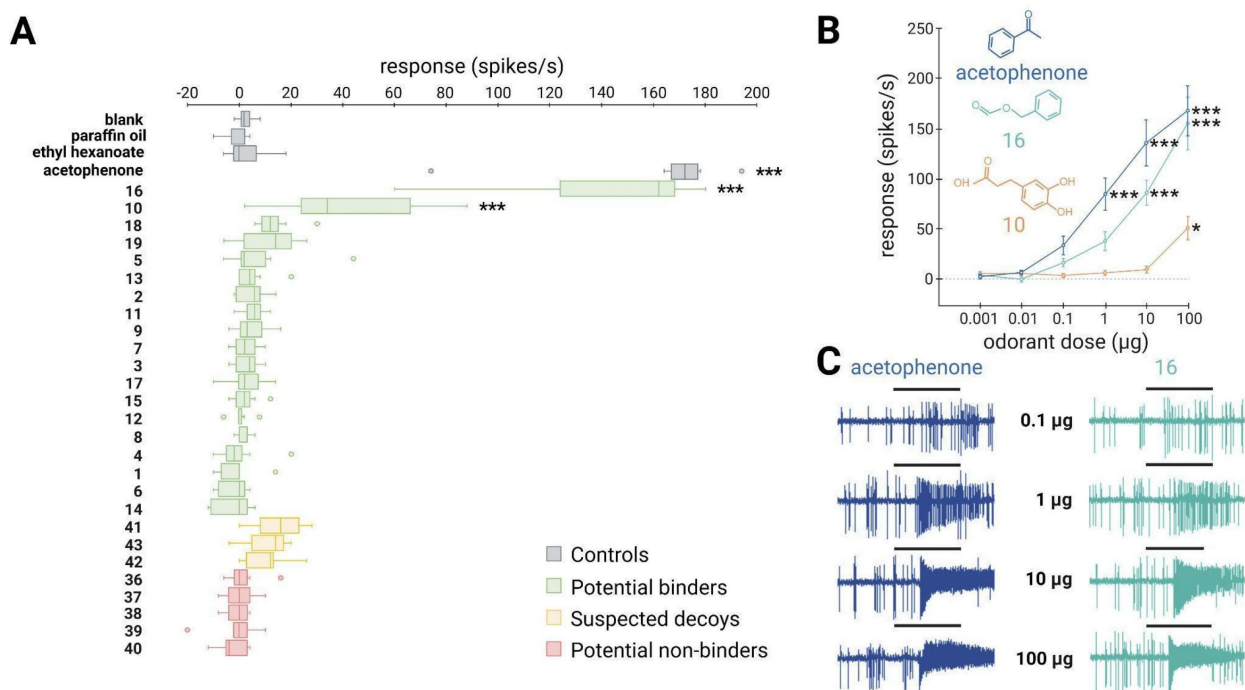
containing SlitOR25 previously identified ligands<sup>16,20</sup>. Remarkably, 3-(3,4-dihydroxyphenyl)propanoic acid (10) was part of a new cluster (Fig. 3 and Figure S3). By comparing this new ligand with known ligands, we found that the maximum Tanimoto coefficient obtained is low ( $T_c=0.25$ ), revealing unusual chemical structures for a SlitOR25 ligand. Dose-response experiments were conducted for the two new SlitOR25 ligands discovered in this study (Fig. 4B-C). Compound 10 showed significant receptor activation only at 100  $\mu\text{g}$  on the filter paper. SlitOR25 appeared to be more sensitive to compounds 16, which significantly activated the receptor starting from 10  $\mu\text{g}$ . However, one has to keep in mind that the differences in volatility between the compounds could impact on the real quantity that reaches the neurons and thus may bias the sensitivity interpretation. Notably, the predicted VP of compound 10 is  $10^6$  times lower than that of compound 16 or acetophenone (Table S4).

Out of the 17 molecules predicted to activate SlitOR31, only 2-methoxy-4-propylphenol (23) induced significant responses, resulting in a success rate of approximately 6% (Fig. 5A). Compounds 23 was part of a cluster that also contained eugenol, the

only known ligand of SlitOR31 (Fig. 3 and Figure S3). In line with the findings for SlitOR25, no significant response was observed for the selected molecules predicted as non-binders or suspected decoys. To delve deeper into SlitOR31 sensitivity to compound 23, we conducted dose-response experiments (Fig. 5B-C). This compound significantly activated SlitOR31, with activity observed starting from 10  $\mu\text{g}$ . Despite having a VP similar to that of compound 23 (Table S4), eugenol activated SlitOR31 at 1  $\mu\text{g}$ , reaffirming its position as the most potent ligand for this receptor.

### The binding sites of SlitOR25 and SlitOR31 differ both in their amino acid composition and interaction patterns

To further understand why the broadly tuned receptor SlitOR25 and the specific receptor SlitOR31 exhibited divergent chemical spaces, we thoroughly explored the conservation of their binding site and the interactions of all compounds identified as ligands from this and previous studies<sup>16,20</sup>. Data from 33 molecules for SlitOR25 and two for SlitOR31 were compiled. While residues forming the pore of the ion channel in the OR/Orco complex were significantly



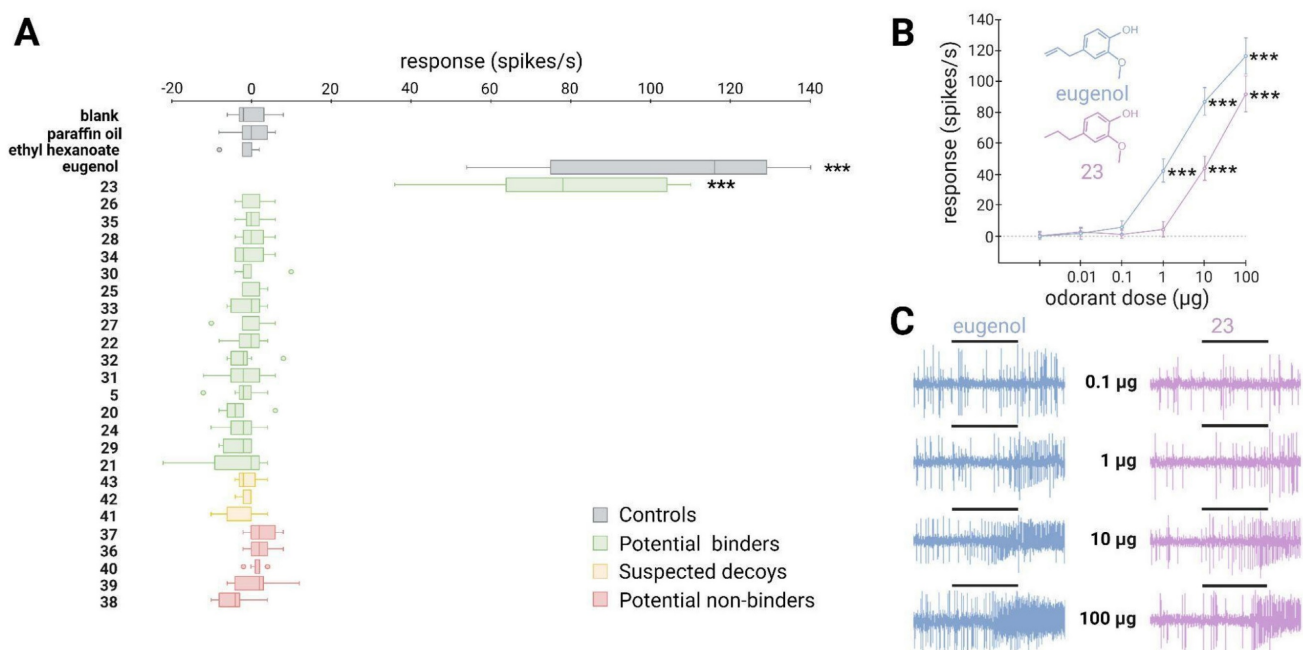
**Figure 4:** ~10% of the predicted binders via the SBVS approach activate the broadly-tuned receptor SlitOR25. (A) Box plot shows the responses of *Drosophila* ab3A OSNs ( $n = 7$ ) expressing SlitOR25 measured upon exposure to a panel of predicted binders, decoys, and non-binders (100  $\mu\text{g}$  loaded in the stimulus cartridge). The whiskers were extended to 1.5 times the interquartile range from the quartiles. Outliers are indicated by dots. The controls are represented in grey (blank, paraffin oil solvent, ethyl hexanoate and acetophenone). Green boxes represent the responses to predicted binder compounds, yellow boxes depict the responses to predicted binder compounds that have been removed from our selection because the majority of the tested SlitORs (suspected decoys), and red boxes indicate the responses to predicted non-binder compounds. (B) Dose-response curves of *Drosophila* ab3A OSNs for all active compounds revealed during the screening of the panel. Data represented are mean action potential frequencies  $\pm$  SEM ( $n = 7$ ). In (A) and (B), asterisks indicate statistically significant differences between responses to the odorant and to the solvent (Friedman test followed by a Dunnett's multiple comparison test with a Benjamini & Hochberg correction; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (C) Examples of single-sensillum recording traces obtained for *Drosophila* ab3A OSNs expressing SlitOR25 stimulated with increasing doses of acetophenone (known ligand) and benzyl formate (16, new active compound which exhibited the highest response). Black bars represent the duration of the stimulus (500 ms).

conserved within the SlitORs, the residues forming the binding sites exhibited higher variability (Fig. 6A-C). These differences could potentially correlate with the diverse chemical spaces detected by the SlitORs.

SlitOR25 ligands appeared to bind mainly via hydrophobic interactions, and following a non-conserved ligand-receptor interacting scheme. The following thirteen residues interacted with at least one ligand: V88, L91, Q92, T153, L154, W157, F197, I200, S201, V204, M205, F322, and Y325. As shown in Fig. 6A-C, a majority of active compounds were also engaged in a unique polar interaction mainly with T153 and/or S201. The new ligand (10) is a notable exception, exhibiting three hydrogen bond interactions. It is reasonable to classify these polar interactions as hydrogen bonds even if some do not fully meet all structural criteria for conventional hydrogen bonds and should be considered as weak hydrogen bonds. Interestingly, SlitOR25 activation appeared to be versatile, as it did not strictly adhere to a predefined interaction pattern. Six residues (V88, L91, L154, W157, F197, and F322) constituted the core of the binding site and exhibited hydrophobic interactions with over 90% of the known ligands. Among them, W157 and F322, which form the

binding pocket's lid, were the most highly conserved residues in the binding pocket across SlitORs (Fig. 6C). The aromatic amino acid W157 aligned with W158 in the MhraOR5 structure<sup>3</sup> and with F115 in the ApisOR5 structure<sup>5</sup> which are key residues for ligand detection. Moreover, residue W157 also aligned with F136 in the AegOR10 structure<sup>6</sup> where its rotation enables the transition between the inactive and active states of the receptor. Conversely, residues Q92, I200, S201, V204, M205 and Y325 displayed hydrophobic interactions with less than 46% of the ligands, suggesting that the binding pocket has the capacity to generate additional anchoring points for various odorant structures, in line with the receptor's wide range of recognition capabilities.

Looking at SlitOR31, 10 residues (F78, I79, T82, I141, C144, G145, W148, I189, A193 and Y313) were found to interact with the ligands. The ligand binding was mediated by hydrophobic and polar interactions that were widely shared across known ligands (Fig. 6B, C). Some structural features of the binding pocket were common between SlitOR31 and SlitOR25. In each case, the ligands were stabilized through hydrophobic interactions with a subset of aliphatic and aromatic side chains. As for SlitOR25, W158 and Y313, which are among the most conserved residues



**Figure 5: A new binder predicted via the SBVS approach activates the narrowly-tuned receptor SlitOR31.** (A) Box plot shows the responses of *Drosophila* ab3A OSNs ( $n = 7$ ) expressing SlitOR31 measured upon exposure to a panel of predicted binders, decoys, and non-binders (100  $\mu\text{g}$  loaded in the stimulus cartridge). The whiskers were extended to 1.5 times the interquartile range from the quartiles. Outliers are indicated by dots. The controls are represented in grey (blank, paraffin oil solvent, ethyl hexanoate and eugenol). Green boxes represent the responses to predicted binder compounds, yellow boxes depict the responses to predicted binder compounds that have been removed from our selection because they activate the majority of the tested SlitORs, and red boxes indicate the responses to predicted non-binder compounds. (B) Dose-response curves of *Drosophila* ab3A OSNs for all active compounds revealed during the screening of the panel. Data represented are mean action potential frequencies  $\pm$  SEM ( $n = 7$ ). In (A) and (B), asterisks indicate statistically significant differences between responses to the odorant and to the solvent (Friedman test followed by a Dunnett's multiple comparison test with a Benjamini & Hochberg correction; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ). (C) Example of single-sensillum recording traces obtained for *Drosophila* ab3A OSNs stimulated with increasing doses of eugenol (known ligand) and 2-methoxy-4-propylphenol (23, new active compound). Black bars represent the duration of the stimulus (500 ms).

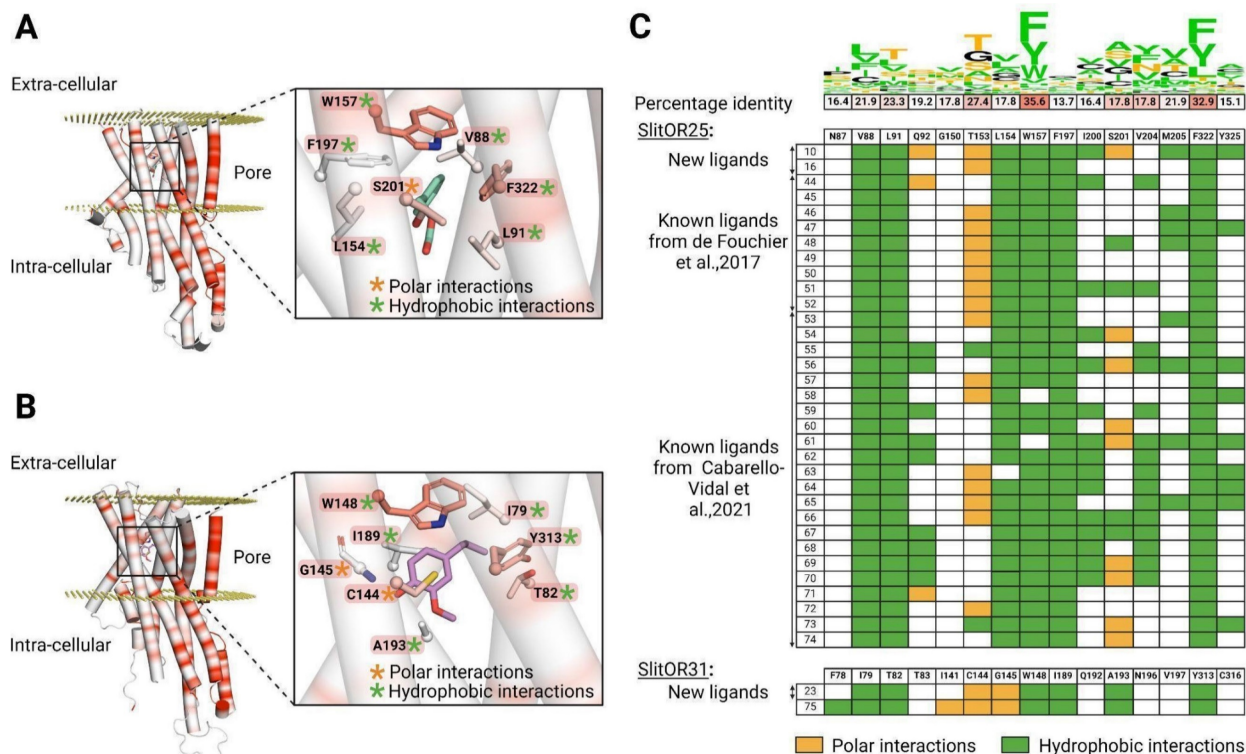
across SlitOR binding pockets, were observed to interact with every ligand and contribute to form the binding pocket's lid. Here again, polar interactions might involve either typical hydrogen bonds or weaker variations, such as the polarized S-H bond in the C144 thiol group, which exhibits reduced polarity compared to a standard hydrogen bond. In contrast, SlitOR25 residues interact via their side chains, while interactions with SlitOR31 residues I141 and G145 occurred through backbone atoms and polar functional groups of the odorants.

In total, ten residues of the SlitOR25 binding pocket and nine residues of the SlitOR31 binding pocket were aligned with amino acids in the binding pockets of MhraOR5<sup>3</sup>, ApisOR5<sup>5</sup>, AegOR10<sup>6</sup> or DmelOrco<sup>66</sup>, where mutations have been shown to significantly impact ligand-receptor interactions (Table S5). Therefore, we hypothesize that these residues play a critical role in the activation of these two receptors.

### The two new ligands of SlitOR25 are highly attractive to *S. littoralis* larvae

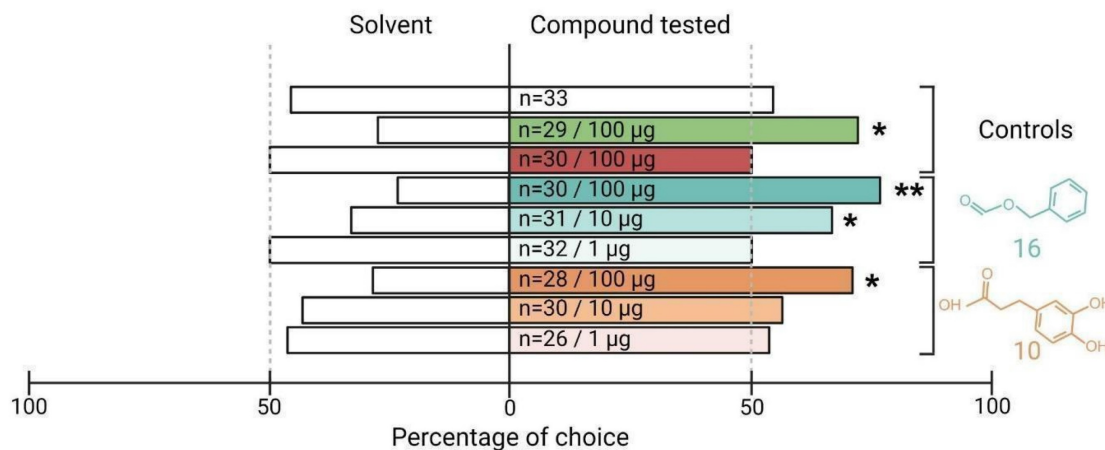
The electrophysiological tests have confirmed or refuted the predicted activity of potential binders identified by our SBVS workflow on SlitOR25 and SlitOR31. To go further, we next investigated the biological relevance of the newly identified ligands on *S. littoralis* behavior using a dual-choice assay in a Y-tube olfactometer. Since previous studies have

demonstrated that SlitOR25 activation leads to larvae attraction<sup>16,62</sup>, we speculated that the new ligands would induce a similar behavior. By contrast, the primary ligand of SlitOR31, eugenol, being not known to induce any specific larvae behavior<sup>62</sup>, we did not investigate the effect of the new SlitOR31 ligand. The viability of our experimental protocol was verified using three control compounds: paraffin oil (solvent), benzyl alcohol (known attractant), and (E)-ocimene (known to be neutral). All controls resulted in the expected larval responses: no choice for a specific arm in paraffin oil vs paraffin oil and paraffin oil vs (E)-ocimene assays, choice to the odorized arm in paraffin oil vs benzyl alcohol assay (Fig. 7). We next tested the effects of the new SlitOR25 ligands, benzyl formate (16) and 3-(3,4-dihydroxyphenyl)propanoic acid (10) vs paraffin oil on larval choice, and at different doses : 100, 10, and 1  $\mu$ g. At 100  $\mu$ g, 76.7% and 71.4% of the larvae made a choice to the arm odorized with compounds 16 and 10, respectively. Compound 16 retained activity at 10  $\mu$ g, with a choice percentage of 67.7, but not at 1  $\mu$ g, whereas compound 10 did not induce any choice at 1 and 10  $\mu$ g. These results correlate with the difference we observed in SlitOR25 sensitivity to the two compounds during the single-sensillum recordings, where the detection threshold of SlitOR25 was lower for benzyl formate (16) than for 3-(3,4-dihydroxyphenyl)propanoic acid (10).



**Figure 6: Activation patterns differ between a broadly-tuned receptor (SlitOR25) and a specific receptor (SlitOR31).** (A) Structure of SlitOR25 and (B) SlitOR31 respectively interacting with benzyl formate (16) and 2-methoxy-4-propylphenol (23), the newly discovered ligands inducing the most robust responses from their

respective receptor. The receptor orientation in the membrane was determined using OPM server<sup>65</sup>. Orange and green asterisks symbolize polar and hydrophobic interactions, respectively, formed between the ligand and residues within the receptor's binding pocket. Additionally, a gradient color coding across the entire protein highlights the conservation of amino acids among the 73 odorant receptors of *S. littoralis*: highly conserved residues appearing in red and non-conserved ones in white. The backbones of the residues are shown only if they interact with the molecules; otherwise, only the side chains are visualized. This figure was generated using the molecular visualization software PyMol<sup>69</sup>. (C) Comprehensive overview of ligand-receptor interaction pattern across all known natural compounds activating SlitOR25 and SlitOR31. We utilize the same color code as used in (A) and (B): hydrophobic interactions are depicted in green and polar interactions in orange. At the top of this table, the percentage of identity among interacting residues within the 73 SlitORs is indicated and coupled with a Show Logo. Hydrophobic residues are highlighted in green. Residues that can potentially form polar interactions are highlighted in orange, except for Y and W that are already colored in green. The others are in black.



**Figure 7: The two new ligands of SlitOR25 discovered through our Structure-Based Virtual Screening workflow are attractive for the larvae of *S. littoralis*.** Data are presented as a percentage of larval choice. Controls: blank control (paraffin oil/paraffin oil, white bar), positive control (paraffin oil/benzyl alcohol, green bar), neutral control (paraffin oil/(E)-ocimene, red bar). The choices (paraffin oil/SlitOR25 new ligands) related to benzyl formate (16) and 3-(3,4-dihydroxyphenyl)propanoic acid (10) are represented in cyan and orange, respectively. The tested doses and the number of replicates (n) are indicated inside each bar of the diagram. Asterisks indicate statistically significant preferences of larvae for the odorized arm (Chi-squared test; \*p < 0.05, \*\*p < 0.01).

## Discussion

A better understanding of an animal chemical ecology requires extensive knowledge on the odorants it perceives and the resulting effects. Traditionally, chemical ecology approaches examine naturally occurring chemically-mediated processes in the species environment, focusing on known food diet, animal, and environmental volatiles ecologically relevant to the targeted species, following an *a priori* strategy. On the contrary, reverse chemical ecology investigates the underlying mechanisms to identify the proteins involved (e.g. odorant receptors), to be used as targets for large screens without any *a priori* on ecological relevance, offering the possibility to explore new areas of chemical spaces. In that view, investigating novel chemicals targeting insect ORs holds considerable promise for expanding our fundamental understanding of their chemical ecology and enabling practical advancements in pest control *via* the development of new behavioral disruptors. In the present work, we set up a joint computational and experimental approach to extend the chemical spaces of ORs from the crop pest moth *Spodoptera littoralis*. In particular, we focused on one broadly and one narrowly tuned OR to virtually screen a large database of natural compounds, coupled to experimental validation of the predicted ligands and behavioral assays. Unlike ligand-based methods which rely on prior knowledge of ligand structures,

the Structure-Based Virtual Screening pipeline we propose here efficiently processed large libraries and utilized predictive modeling that incorporated established molecular recognition principles to discover compounds in both explored and unexplored chemical spaces. To our knowledge, this is the first report of a SBVS approach successfully employed for insect ORs. With a success rate ranging from 6% to 10% in accurately predicting active compounds, we demonstrate the power of the approach. As underlined in a recent review dedicated to docking simulations, a hit rate of 5% or even 3% is considered more than acceptable<sup>67</sup>. This arises when limited information about the target is available or when the screening library has not undergone any preliminary filtering. Higher performance can be reached if an experimental structure is available or, for instance, the location of the binding site or key interacting residues are known. Moreover, our approach not only identified potential binders but also effectively distinguished non-binders.

However, the SBVS approach is far from perfect and fails in some cases. It may be due to discrepancies in the AlphaFold2-predicted structure of the binding pocket, as has recently been highlighted<sup>68,69</sup>. Comparing the models we obtained using AlphaFold2 with those obtained through the newly released version AlphaFold3<sup>47</sup>, which has been specifically designed for addressing this issue, revealed no significant improvement. The

Root-Mean-Square Deviation (RMSD) between AlphaFold2 and AlphaFold3 models was always lower than 1.2 Å for the entire proteins and lower than 0.4 Å on the binding pocket (Figure S1A-B). Moreover, the performance of AlphaFold2 and AlphaFold3 in predicting the structures of AegOR10, AgamOR28, ApisOR5, MhraOR5, and AbakOrco was comparable (Table S6). Due to the high divergence shown by insect ORs, TBM could outperform AlphaFold for some species, such as *Bactrocera*<sup>70</sup>. Homology models generated by tools like MODELLER<sup>48</sup> could then be considered an alternative to AlphaFold in our workflow. To test this, we conducted blind predictions of the experimental structures of insect ORs using either AlphaFold2, AlphaFold3, or MODELLER (Table S6). Here, sequence identity between SlitORs and templates is low, ranging from 12.9 to 21.2% (Table S7) and it appeared that AlphaFold2 in our workflow was the most reliable option to ensure accurate receptor modeling (Table S6). These findings are in agreement with Lee et al.<sup>71</sup> showing that MODELLER is generally more accurate when the query sequence closely matches the template while AlphaFold tends to outperform homology modeling, particularly when the protein lacks a closely related experimental structure. In addition, it would also have been possible to explore the potential gains from considering flexible receptor docking for the experimental dataset consisting of 51 molecules. However, this approach would have significantly increased both computation time and resource requirements for the database containing several hundred thousand molecules. Indeed, the current trend is not to complexify the docking algorithm but rather to apply it on a subset of a large library by prioritizing molecules that occupy the same chemical space as potential hits (by calculating chemical similarity and clustering)<sup>72,73</sup> or by training a machine learning model capable of screening the entire database<sup>74,75</sup>. We can then consider that the scoring function alone is not sufficient to accurately rank and distinguish binders and non-binders. As shown in Fig. 2, the Vinardo scoring function performed well for pheromone receptors and poorly for non-pheromonal ORs, which led us to recalibrate the scoring function. It should be noted that the formation of a ligand-receptor complex is driven by the free energy of binding, which can be decomposed into two components: an enthalpic and an entropic contribution. In docking scoring functions, the entropic component may not be accurately accounted for or even overlooked entirely, often being approximated solely based on the number of rotatable bonds within the molecule. As a result, large molecules with high molecular weights are typically

able to fit better into the binding site and to establish more interactions with the target protein, resulting in higher docking scores. This is what we suspected by comparing docking scores with *in vivo* responses for various classes of ligands (Figure S4). A potential solution to the overestimation of the enthalpic term (or an underestimation of the entropic penalty) was to adjust the scoring function so that it takes into consideration the molecular weight (or the number of heavy atoms) of each compound. We tested various rescoring methods and observed that weighting, such as the Ligand Efficiency (LE) score, provides the best performance. It is not so usual to replace the docking scoring function by LE but it has already been useful in previous studies<sup>52,53,54</sup>. However, using Ligand Efficiency scoring reversed the tendency and wrongly ranked very small molecules as top ligands. This suggests that solving the problem would require the optimization of a new scoring function including a parameter accounting for molecular weight. Moreover, the rescoring did not have the same effect on all the receptors. For instance, when considering SlitOR3, the improvement was modest, as shown in Fig. 2A (AUC increases from 0.52 to 0.59 when using LE instead of Vinardo). A possible rationale lies in the application of a rigorous statistical criterion (*p*-value below 0.001) for categorizing compounds as either binders or non-binders. Adopting a less stringent *p*-value threshold enhanced docking performance, as demonstrated in Figure S5. This observation underscores the significance of carefully curating databases tailored for building meaningful computational models. So, to avoid a bias in the optimization function due to a limited dataset of 850 ligand-receptor pairs (51 odorants tested with 17 SlitORs), we finally decided to define a specific filter on the docking output results to exclude molecules displaying top rank position across nearly all receptors as it might actually be attributed to a scoring error, rather than depicting true biological activity. Defining such a “suspected decoy” filter effectively reduced the number of false positives and may be a viable compromise to maximize docking performance without the need to create a new scoring function from scratch.

Beyond its ability to explore new chemical spaces, the strength of the SBVS approach lies in its capability to unravel structure-function relationships, which sets it apart from LBVS. For instance, SBVS can guide site-directed mutagenesis for investigating the activation mechanisms of ORs. The analysis of SlitOR25 and SlitOR31 binding cavities highlighted key residues, aligned with those found in the experimental MhraOR5<sup>3</sup>, ApisOR5<sup>5</sup>, AegOR10<sup>6</sup>, and DmelOrco<sup>66</sup> structures (Table S5), that could be critical

for the ligand binding and the activation mechanism. Moreover, the analysis of ligand-receptor interactions shows that in most of the cases SlitOR25 forms fewer hydrogen bonds with ligands than SlitOR31. This difference correlated with the higher polarity scores for the binding site of SlitOR31 compared to SlitOR25. Additionally, the cavity volume of SlitOR31 slightly exceeded that of SlitOR25 (roughly 530 vs 300 Å<sup>3</sup>). It is consistent with the average number of heavy atoms (12 vs 8.4±1.4) and average size of SlitOR31 and SlitOR25 ligands (165±4 vs 123±13 Å<sup>3</sup>). Our hypothesis is that the recognition spectrum of the receptor is partly driven by the size and polarity of the binding site. To support our findings (especially because of the limited number of SlitOR31 ligands), we analyzed the binding pockets of the 17 SlitORs from de Fouchier *et al.*<sup>20</sup> Since all of these receptors were tested with the same 51 odorants, this provides clues about their tuning, enabling us to elaborate on structure-function relationship hypotheses. As shown in Figure S6 and Table S8, the size and hydrophobicity of the binding pocket are correlated to the number of active molecules per receptor. Even if the correlation is somehow modest (Pearson's  $r \sim 0.4-0.6$ ), the larger the volume (or the solvent accessible surface area - SASA - of the pocket) and the lower the hydrophobicity, the narrower is the recognition spectrum of the receptor. This observation on the SlitORs aligns with findings from the comparison of SlitOR31 and MhraOR5, suggesting that such a tendency might be extrapolated to other insect ORs. Despite both receptors are able to detect eugenol, no specific amino acid pattern associated with eugenol detection was found, as shown in Figure S7B. In contrast, the differences between broad and specific receptors, as noted earlier, remain apparent. MhraOR5, a broadly tuned receptor, recognizes eugenol primarily through hydrophobic interactions, while SlitOR31 recognizes eugenol through a mechanism that relies more heavily on polar interactions, as shown in Figure S7A. We also attempted to build a simple model to predict the recognition spectrum of SlitORs. We set up a multiple linear regression model based on different combinations of the binding pocket descriptors to predict the experimental number of active molecules. As shown in Figure S8 and Table S9, a very simple model with only two variables, SASA and hydrophobicity, performed well with a correlation coefficient of 0.75 between the experimental and the predicted values and a Root Mean Square Error (RMSE) of 2.53. It already captured a large part of the structure-function relationship and was not much less powerful than the model with a higher number of variables (with 6 pocket descriptors,  $r=0.79$  and

RMSE=2.35). However, these conclusions must be put into perspective in view of the small size of the dataset. Moreover, a recent study on an OR-related sugar receptor highlights that chemical specificity is not exclusively determined by the selectivity of the ligand-binding pocket, but rather derives from multiple factors, primarily receptor-ligand interactions and allosteric coupling<sup>76</sup>. Anyhow, such a binding pocket analysis could guide the prediction of orphan receptors' broadness.

The two new SlitOR25 ligands we identified in this study were attractants for *S. littoralis* larvae, as were the previously identified ligands<sup>16,62</sup>. In addition to reaffirming the effectiveness of reverse chemical ecology approaches, this discovery supports the hypothesis that activation of SlitOR25 induces larval attraction in *S. littoralis*. To our knowledge, the two new ligands have never been investigated in *S. littoralis* chemical ecology studies, and their ecological relevance can be questioned. Benzyl formate (16) induced the highest firing rate in OSNs expressing SlitOR25 and demonstrated a strong attraction rate for the larvae. Interestingly, this compound has been reported in volatile emissions of plant species within the genus *Solanum*: *S. stuckertii* and *S. incisum*<sup>77</sup>. As this plant genus includes the major host plants of *S. littoralis*, such as *S. lycopersicum*, *S. melongena* and *S. tuberosum*<sup>78</sup>, this cue may indicate the presence of suitable food sources for the larvae if emitted as well by these latter species. 3-(3,4-dihydroxyphenyl) propanoic acid (10) was the second compound we identified as activating SlitOR25 and inducing attraction of *S. littoralis* larvae. This molecule is present in various plant species, spanning herbs, shrubs and trees from distantly related families<sup>79</sup>. Interestingly, none of these plant species are known hosts of *S. littoralis*<sup>78</sup>, although some are present in the insect's natural environment. As plant volatiles are constituted by a diversity of molecules, it is possible that other volatiles than 3-(3,4-dihydroxyphenyl) propanoic, detected by other SlitORs than OR25, disrupt the larvae attraction to this compound, leading to avoidance. Alternatively, unpalatable molecules may refrain larvae from feeding on these plants. The behavioral effect of the newly discovered ligand for SlitOR31, 2-methoxy-4-propylphenol (23), has not been tested in the current study since eugenol (the only previously identified ligand) is known to be inactive on *S. littoralis* larvae behavior<sup>62</sup>. However, eugenol has been shown to have antifeedant activity in larvae, resulting in their death if no other food source is present<sup>80</sup>. The close structural similarity between eugenol and 2-methoxy-4-propylphenol, only differing by the presence or absence of a double bond at the end of the aliphatic chain, suggests that

both compounds may have similar effects on *S. littoralis*. This hypothesis is supported by a study demonstrating that applying 100 µg of 2-methoxy-4-propylphenol to a piece of maize significantly reduces feeding activities of larvae from the related species *S. frugiperda*<sup>81</sup>. A literature survey identified this compound in plant defense emissions of the oak *Quercus agrifolia*<sup>82</sup>. Whether it is also produced in non-host plants of *S. littoralis* remains to be investigated. Avoiding these cues could indeed enhance larval survival chances.

## Conclusions

Our study paves the way for future research aimed at expanding the insect OR chemical space, which defines a given species' olfactory capacities and shapes its chemical ecology. A comprehensive analysis of complete OR chemical spaces in both closely and distantly related species will ultimately unravel the intricate mechanisms of OR functionality and specialization and help us understand how OR evolution contributes to species adaptation to specific ecological niches. The Structure-Based Virtual Screening approach described here also opens new avenues for pest control via olfactory disruption, as it will undoubtedly accelerate the discovery of new behaviorally active volatiles.

## Abbreviations

**ORs:** Odorant receptors; **SBVS:** Structure-Based Virtual Screening; **TM:** Transmembrane domain; **OSNs:** Olfactory sensory neurons; **GPCR:** G protein-coupled receptor; **Orco:** Odorant receptor co-receptor; **Cryo-EM:** Cryo-electron microscopy; **LBVS:** Ligand-Based Virtual Screening; **OBPs:** odorant-binding proteins; **Slit:** *Spodoptera littoralis*; **Aaeg:** *Aedes aegypti*; **Agam:** *Anopheles gambiae*; **Apis:** *Acyrtosiphon pisum*; **Mhra:** *Machilis hrabei*; **Abak:** *Apocrypta bakeri*; **TPR:** True positive rate; **FPR:** False positive rate; **Dmel:** *Drosophila melanogaster*; **ROC:** Receiver operating characteristic; **AUC:** Area under the ROC curve; **PRs:** Pheromone receptors; **VP:** Predicted vapor pressure at 25°C; **RMSD:** The Root-Mean-Square Deviation; **SASA:** Solvent Accessible Surface Area; **RMSE:** Root Mean Square Error.

## Supplementary Material

Supplementary figures and tables 2-9.

<https://www.ijbs.com/v21p2101s1.pdf>

Supplementary table 1.

<https://www.ijbs.com/v21p2101s2.xlsx>

## Acknowledgements

The authors thank Thomas Chertemps for insightful comments and fruitful discussions. This project has received financial support from the CNRS through the 80 | Prime program, the French National Research Agency (ANR-20-CE20-003), and the French government through the France 2030 investment plan managed by the National Research Agency (ANR) through the ExplorAE program, and as part of the Initiative of Excellence Université Côte d'Azur under reference number ANR-15-IDEX-01. The authors are grateful to the Université Côte d'Azur's Center for High-Performance Computing (OPAL infrastructure) for providing resources and support. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

E.J.J and S.F. designed research; A.C, M.L., L.B. and R.M. performed research; A.C. analyzed data; A.C., E.J.J. and S.F. wrote the paper. All authors discussed the results, reviewed and approved the manuscript.

## Competing Interests

The authors have declared that no competing interest exists.

## References

- Montagné N, Gévar J, Lucas P. Semiochemicals and communication in insects. In: Extended Biocontrol, ed. Springer; 2022: 173-181. [https://doi.org/10.1007/978-94-024-2150-7\\_15](https://doi.org/10.1007/978-94-024-2150-7_15)
- Butterwick JA, del Marmol J, Kim KH, et al. Cryo-EM structure of the insect olfactory receptor Orco. Nature. 2018; 560: 447-452. <https://doi.org/10.1038/s41586-018-0420-8>
- Del Marmol J, Yedlin MA, Ruta V. The structural basis of odorant recognition in insect olfactory receptors. Nature. 2021; 597: 126-131. <https://doi.org/10.1038/s41586-021-03794-8>
- Sato K, Pellegrino M, Nakagawa T, et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature. 2008; 452: 1002-1006. <https://doi.org/10.1038/nature06850>
- Wang Y, Qiu L, Wang B, et al. Structural basis for odorant recognition of the insect odorant receptor OR-Orco heterocomplex. Science. 2024; 384: 1453-1460. <https://doi.org/10.1126/science.adn6881>
- Zhao J, Chen AQ, Ryu J, Del Marmol J. Structural basis of odor sensing by insect heteromeric odorant receptors. Science. 2024; 384: 1460-1467. <https://doi.org/10.1126/science.adn6384>
- Benton R, Dessimoz C, Moi D. A putative origin of the insect chemosensory receptor superfamily in the last common eukaryotic ancestor. Elife. 2020; 9: e62507. <https://doi.org/10.7554/eLife.62507>
- Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. Neuron. 2004; 43: 703-714. <https://doi.org/10.1016/j.neuron.2004.08.019>
- Wicher D, Schäfer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, et al. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature. 2008; 452: 1007-1011. <https://doi.org/10.1038/nature06861>
- Bai PH, Wang B, Zhang XH, Wang GR. Research methods and advances of odorant receptors in insects. Acta Entomologica Sinica. 2022; 65: 364-385. <https://doi.org/10.16380/j.kcxb.2022.03.012>
- Ioannidis P, Simao FA, Waterhouse RM, Manni M, Seppey M, Robertson HM, et al. Genomic features of the damselfly *Calopteryx splendens* representing a sister clade to most insect orders. Genome biology and evolution. 2017; 9: 415-430. <https://doi.org/10.1093/gbe/evx006>
- Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, Ray A, et al. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant

- species reveals sex-specific signatures of odor coding. *PLoS Genetics*. 2012; 8: e1002930. <https://doi.org/10.1371/journal.pgen.1002930>
- [13] Robertson HM. Molecular evolution of the major arthropod chemoreceptor gene families. *Annual review of entomology*. 2019; 64: 227-242. <https://doi.org/10.1146/annurev-ento-020117-043322>
- [14] Boyle SM, McInally S, Ray A. Expanding the olfactory code by in silico decoding of odor-receptor chemical space. *Elife*. 2013; 2: e01120. <https://doi.org/10.7554/eLife.01120>
- [15] Caballero-Vidal G, Bouysset C, Grunig H, Fiorucci S, Montagné N, Golebiowski J, et al. Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor. *Scientific Reports*. 2020; 10: 1655. <https://doi.org/10.1038/s41598-020-58564-9>
- [16] Caballero-Vidal G, Bouysset C, Gévar J, Mbouzid H, Nara C, Delaroche J, et al. Reverse chemical ecology in a moth: machine learning on odorant receptors identifies new behaviorally active agonists. *Cellular and Molecular Life Sciences*. 2021; 78: 6593-6603. <https://doi.org/10.1007/s00018-021-03919-2>
- [17] Chahda JS, Soni N, Ebrahim S, Weiss BL, Carlson JR. The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLoS Genetics*. 2019; 15: e1008005. <https://doi.org/10.1371/journal.pgen.1008005>
- [18] Kepchia D, Xu P, Terryn R, Castro A, Schürer SC, Leal WS, et al. Use of machine learning to identify novel, behaviorally active antagonists of the insect odorant receptor co-receptor (Orco) subunit. *Scientific Reports*. 2019; 9: 4055. <https://doi.org/10.1038/s41598-019-40640-4>
- [19] Oliferenko PV, Oliferenko AA, Poda GI, et al. Promising *Aedes aegypti* repellent chemotypes identified through integrated QSAR, virtual screening, synthesis, and bioassay. *PLoS One*. 2013; 8: e64547. <https://doi.org/10.1371/journal.pone.0064547>
- [20] De Fouchier A, Walker III WB, Montagné N, et al. Functional evolution of Lepidoptera olfactory receptors revealed by deorphanization of a moth repertoire. *Nature communications*. 2017; 8: 15709. <https://doi.org/10.1038/ncomms15709>
- [21] Pask GM, Slone JD, Millar JG, et al. Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. *Nature communications*. 2017; 8: 297. <https://doi.org/10.1038/s41467-017-00099-1>
- [22] Jumper J, Evans R, Green T, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021; 596: 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- [23] Ballante F, Kooistra AJ, Kampen S, de Graaf C, Carlsson J. Structure-based virtual screening for ligands of G protein-coupled receptors: what can molecular docking do for you? *Pharmacological Reviews*. 2021; 73: 1698-1736. <https://doi.org/10.1124/pharmrev.120.000246>
- [24] Zhang M, Chen T, Lu X, et al. G protein-coupled receptors (GPCRs): advances in structures, mechanisms, and drug discovery. *Signal Transduction and Targeted Therapy*. 2024; 9: 88. <https://doi.org/10.1038/s41392-024-01803-6>
- [25] Li Z, Capoduro R, Bastin-Héline L, et al. A tale of two copies: Evolutionary trajectories of moth pheromone receptors. *Proceedings of the National Academy of Sciences*. 2023; 120: e2221166120. <https://doi.org/10.1073/pnas.2221166120>
- [26] Ma Y, Yang TT, Ni S, et al. The Odorant Receptor Recognizing Camphor in a Camphor Tree Specialist *Orthaga achatina* (Lepidoptera: Pyralidae). *Journal of Agricultural and Food Chemistry*. 2024; 72: 2689-2696. <https://doi.org/10.1021/acs.jafc.3c08877>
- [27] Wang JX, Wei ZQ, Chen MD, et al. Conserved odorant receptors involved in nonanal-induced female attractive behavior in two Spodoptera Species. *Journal of Agricultural and Food Chemistry*. 2023; 71: 13795-13804. <https://doi.org/10.1021/acs.jafc.3c03265>
- [28] Xu L, Jiang HB, Yu JL, et al. Two odorant receptors regulate 1-octen-3-ol induced oviposition behavior in the oriental fruit fly. *Communications Biology*. 2023; 6: 176. <https://doi.org/10.1038/s42003-023-04551-5>
- [29] Yuvaraj JK, Roberts RE, Sonntag Y, et al. Putative ligand binding sites of two functionally characterized bark beetle odorant receptors. *BMC biology*. 2021; 19: 1-21. <https://doi.org/10.1186/s12915-020-00946-6>
- [30] Sims C, Withall DM, Oldham N, Stockman R, Birkett M. Computational investigation of aphid odorant receptor structure and binding function. *Journal of Biomolecular Structure and Dynamics*. 2023; 41: 3647-3658. <https://doi.org/10.1080/07391102.2022.2053743>
- [31] Tiwari V, Sowdhamini R. Structure modelling of odorant receptor from *Aedes aegypti* and identification of potential repellent molecules. *Computational and Structural Biotechnology Journal*. 2023; 21: 2204-2214. <https://doi.org/10.1016/j.csbj.2023.03.005>
- [32] Liggri PG, Pérez-Garrido A, Tsitsanou KE, et al. 2D finger-printing and molecular docking studies identified potent mosquito repellents targeting odorant binding protein 1. *Insect Biochemistry and Molecular Biology*. 2023; 157: 103961. <https://doi.org/10.1016/j.ibmb.2023.103961>
- [33] Jayanthi KPD, Kempraj V, Aurade RM, et al. Computational reverse chemical ecology: virtual screening and predicting behaviorally active semiochemicals for *Bactrocera dorsalis*. *Bmc Genomics*. 2014; 15: 1-7. <https://doi.org/10.1186/1471-2164-15-209>
- [34] Thireou T, Kythreoti G, Tsitsanou KE, et al. Identification of novel bioinspired synthetic mosquito repellents by combined ligand-based screening and OBP-structure-based molecular docking. *Insect biochemistry and molecular biology*. 2018; 98: 48-61. <https://doi.org/10.1016/j.ibmb.2018.05.001>
- [35] Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. *Biological Reviews*. 2018; 93: 184-200. <https://doi.org/10.1111/brv.12339>
- [36] Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: collection of open natural products database. *Journal of Cheminformatics*. 2021; 13: 1-13. <https://doi.org/10.1186/s13321-020-00478-9>
- [37] Mayhew EJ, Arayata CJ, Gerkin RC, Lee BK, Magill JM, Snyder LL, et al. Transport features predict if a molecule is odorous. *Proceedings of the National Academy of Sciences*. 2022; 119: e2116576119. <https://doi.org/10.1073/pnas.2116576119>
- [38] [Internet] RDKit: Open-source cheminformatics. <https://www.rdkit.org>
- [39] Ropp PJ, Spiegel JO, Walker JL, Green H, Morales GA, Milliken KA, et al. Gypsum-DL: an open-source program for preparing small-molecule libraries for structure-based virtual screening. *Journal of cheminformatics*. 2019; 11: 1-13. <https://doi.org/10.1186/s13321-019-0358-3>
- [40] O'Boyle NM, Branch M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *Journal of cheminformatics*. 2011; 3: 1-14. <https://doi.org/10.1186/1758-2946-3-33>
- [41] [Internet] Dallakyan S. MGLTools (2010). <http://mglttools.scripps.edu/>
- [42] Dolinsky TJ, Czodrowski P, Li H, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*. 2007; 35: W522-W525. <https://doi.org/10.1093/nar/gkm276>
- [43] Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics*. 2005; 61: 704-721. <https://doi.org/10.1002/prot.20660>
- [44] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*. 2009; 10: 1-11. <https://doi.org/10.1186/1471-2105-10-168>
- [45] Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*. 2011; 27: 3276-3285. <https://doi.org/10.1093/bioinformatics/btr550>
- [46] Bouysset C, Fiorucci S. ProLIF: a library to encode molecular interactions as fingerprints. *Journal of Cheminformatics*. 2021; 13: 72. <https://doi.org/10.1186/s13321-021-00548-6>
- [47] Abramson J, Adler J, Dunger J, Evans R, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024; 630: 493-500. <https://doi.org/10.1038/s41586-024-07487-w>
- [48] Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*. 2016; 54: 5-6. <https://doi.org/10.1002/cpbi.3>
- [49] DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*. 2002; 40: 82-92.
- [50] Masters L, Eagon S, Heying M. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. *Journal of Molecular Graphics and Modelling*. 2020; 96: 107532. <https://doi.org/10.1016/j.jmgm.2020.107532>
- [51] Quiroga R, Villarreal MA, Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLoS one*. 2016; 11: e0155183. <https://doi.org/10.1371/journal.pone.0155183>
- [52] Agarwal R, Smith JC. Speed vs accuracy: effect on ligand pose accuracy of varying box size and exhaustiveness in AutoDock vina. *Molecular Informatics*. 2023; 42: 2200188. <https://doi.org/10.1002/minf.202200188>
- [53] Chang CEA, Chen W, Gilson MK. Ligand configurational entropy and protein binding. *Proceedings of the National Academy of Sciences*. 2007; 104: 1534-1539. <https://doi.org/10.1073/pnas.0610494104>
- [54] Cavalluzzi MM, Mangiatordi GF, Nicolotti O, Lentini G. Ligand efficiency metrics in drug discovery: The pros and cons from a practical perspective. *Expert opinion on drug discovery*. 2017; 12: 1087-1104. <https://doi.org/10.1080/17460441.2017.1365056>
- [55] Ke YY, Coumar MS, Shiao HY, Wang WC, et al. Ligand efficiency based approach for efficient virtual screening of compound libraries. *European journal of medicinal chemistry*. 2014; 83: 226-235. <https://doi.org/10.1016/j.ejmech.2014.06.029>
- [56] Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH. The role of ligand efficiency metrics in drug discovery. *Nature reviews Drug discovery*. 2014; 13: 105-121. <https://doi.org/10.1038/nrd4163>
- [57] McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. 2018; *arXiv preprint arXiv:1802.03426*. <https://doi.org/10.48550/arXiv.1802.03426>
- [58] Campello RJ, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: *Pacific-Asia conference on knowledge discovery and data mining*, ed. Springer; 2013: 160-172. [https://doi.org/10.1007/978-3-642-37456-2\\_14](https://doi.org/10.1007/978-3-642-37456-2_14)
- [59] Duffy JB. GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *genesis*. 2002; 34: 1-15. <https://doi.org/10.1002/gene.10150>
- [60] Chahda JS, Soni N, Sun JS, et al. The molecular and cellular basis of olfactory response to tsetse fly attractants. *PLoS genetics*. 2019; 15: e1008005. <https://doi.org/10.1371/journal.pgen.1008005>
- [61] Münch D, Galizia CG. DoOR 2.0-comprehensive mapping of *Drosophila melanogaster* odorant responses. *Scientific reports*. 2016; 6: 21841. <https://doi.org/10.1038/srep21841>
- [62] De Fouchier A, Sun X, Caballero-Vidal G, Travaillard S, Jacquin-Joly E, Montagné N. Behavioral effect of plant volatiles binding to Spodoptera littoralis larval odorant receptors. *Frontiers in Behavioral Neuroscience*. 2018; 12: 264. <https://doi.org/10.3389/fnbeh.2018.00264>



- [63] Poivet E, Rharrabe K, Monsempe C, et al. The use of the sex pheromone as an evolutionary solution to food source selection in caterpillars. *Nature communications*. 2012; 3: 1047. <https://doi.org/10.1038/ncomms2050>
- [64] Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO. Virtual screening workflow development guided by the “receiver operating characteristic” curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of medicinal chemistry*. 2005; 48: 2534-2547. <https://doi.org/10.1021/jm049092j>
- [65] Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research*. 2012; 40: D370-D376. <https://doi.org/10.1093/nar/gkr703>
- [66] Pacalon J, Audic G, Magnat J, et al. Elucidation of the structural basis for ligand binding and translocation in conserved insect odorant receptor co-receptors. *Nature Communications*. 2023; 14: 8182. <https://doi.org/10.1038/s41467-023-44058-5>
- [67] Paggi JM, Pandit A, Dror RO. The Art and Science of Molecular Docking. *Annual Review of Biochemistry*. 2024; 93: 389-410. <https://doi.org/10.1146/annurev-biochem-030222-120000>
- [68] Karelina M, Noh JJ, Dror RO. How accurately can one predict drug binding modes using AlphaFold models? *Elife*. 2023; 12: RP89386. <https://doi.org/10.7554/eLife.89386.2>
- [69] Lyu J, Kopolka N, Gumpfer R, et al. AlphaFold2 structures guide prospective ligand discovery. *Science*. 2024; 384: eadn6354. <https://doi.org/10.1126/science.adn6354>
- [70] Jabeen A, Oakeshott JG, Lee SF, Ranganathan S, Taylor PW. Template-based modeling of insect odorant receptors outperforms AlphaFold3 for ligand binding predictions. *Scientific Reports*. 2024; 14: 29084. <https://doi.org/10.1038/s41598-024-80094-x>
- [71] Lee C, Su BH, Tseng YJ. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Briefings in bioinformatics*. 2022; 23: bbac308. <https://doi.org/10.1093/bib/bbac308>
- [72] Bender BJ, Gahbauer S, Luttens A, et al. A practical guide to large-scale docking. *Nature protocols*. 2021; 16: 4799-4832. <https://doi.org/10.1038/s41596-021-00597-z>
- [73] Lyu J, Wang S, Balius TE, et al. Ultra-large library docking for discovering new chemotypes. *Nature*. 2019; 566: 224-229. <https://doi.org/10.1038/s41586-019-0917-9>
- [74] Gentile F, Yaacoub JC, Gleave J, et al. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols*. 2022; 17: 672-697. <https://doi.org/10.1038/s41596-021-00659-2>
- [75] Gorgulla C. Recent developments in ultralarge and structure-based virtual screening approaches. *Annual Review of Biomedical Data Science*. 2023; 6: 229-258. <https://doi.org/10.1146/annurev-biodatasci-020222-025013>
- [76] Gomes JV, Singh-Bhagania S, Cenci M, et al. The molecular basis of sugar detection by an insect taste receptor. *Nature*. 2024; 629: 228-234. <https://doi.org/10.1038/s41586-024-07255-w>
- [77] Zygadlo JA, Grosso NR. Volatile constituents from the flowers of *Solanum stuckertii* and *S. incisum*. *Journal of Essential Oil Research*. 1997; 9: 111-113. <https://doi.org/10.1080/10412905.1997.9700728>
- [78] [Internet] EPP0 (2024) EPP0 Global Database. <https://gd.eppo.int/taxon/SPODLI/hosts>
- [79] Rutz A, Sorokina M, Galgonek J, et al. The LOTUS initiative for open knowledge management in natural products research. *Elife*. 2022; 11: e70780. <https://doi.org/10.7554/eLife.70780>
- [80] Abdelgaleil SA, Abou-Taleb HK, Al-Nagar NM, Shawir MS. Antifeedant, growth regulatory and biochemical effects of terpenes and phenylpropenes on *Spodoptera littoralis* Boisduval. *International journal of tropical insect science*. 2020; 40: 423-433. <https://doi.org/10.1007/s42690-019-00093-8>
- [81] Vargas-Méndez LY, Sanabria-Flórez PL, Saavedra-Reyes LM, Merchan-Arenas DR, Kouznetsov VV. Bioactivity of semisynthetic eugenol derivatives against *Spodoptera frugiperda* (Lepidoptera: Noctuidae) larvae infesting maize in Colombia. *Saudi journal of biological sciences*. 2019; 26: 1613-1620. <https://doi.org/10.1016/j.sjbs.2018.09.010>
- [82] Ockels FS, Bonello P, McPherson B, Wood DL. Chemical ecology of sudden oak death/ambrosia beetle interactions. In: *Second Science Symposium: The State of Our Knowledge*. 2006; 1: 423.