

## ODORactor: a web server for deciphering olfactory coding

Xinyi Liu<sup>1,†</sup>, Xubo Su<sup>1,†</sup>, Fei Wang<sup>2,†</sup>, Zhimin Huang<sup>1</sup>, Qi Wang<sup>1</sup>, Zhen Li<sup>1</sup>, Ruina Zhang<sup>1</sup>, Lifang Wu<sup>1</sup>, Yi Pan<sup>1</sup>, Yingyi Chen<sup>1</sup>, Hanyi Zhuang<sup>1,†</sup>, Guoqiang Chen<sup>1</sup>, Ting Shi<sup>1</sup> and Jian Zhang<sup>1,\*</sup>

<sup>1</sup>Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of National Ministry of Education, Rui-Jin Hospital, Shanghai Jiao-Tong University School of Medicine, Shanghai, 200025 and <sup>2</sup>Department of Chemistry, Shanghai Institute of Pharmaceutical Industry, Shanghai, 200437, China

Associate Editor: John Quackenbush

### ABSTRACT

**Summary:** ODORactor is an open access web server aimed at providing a platform for identifying odorant receptors (ORs) for small molecules and for browsing existing OR-ligand pairs. It enables the prediction of ORs from the molecular structures of arbitrary chemicals by integrating two individual functionalities: odorant verification and OR recognition. The prediction of the ORs for several odorants was experimentally validated in the study. In addition, ODORactor features a comprehensive repertoire of olfactory information that has been manually curated from literature. Therefore, ODORactor may provide an effective way to decipher olfactory coding and could be a useful server tool for both basic olfaction research in academia and for odorant discovery in industry.

**Availability:** Freely available at <http://mdl.shsmu.edu.cn/ODORactor>

**Contact:** jian.zhang@sjtu.edu.cn

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on February 17, 2011; revised on April 26, 2011; Accepted on June 20, 2011

## 1 INTRODUCTION

The detection of volatile odorants is mediated by odorant receptors (ORs) in the olfactory sensory neurons of the nose (Buck and Axel, 1991). Mammalian ORs comprise the largest subfamily of G protein-coupled receptors on the cell-surface membranes of these neurons. To understand how the plethora of ligands in the environment are detected and discriminated at the molecular level, it is essential to match ORs with their cognant odorant molecules. Previous studies attempted to compile pertinent information of ORs into a database format. The most comprehensive OR database to date is the Olfactory Receptor Database (ORDB), which includes sequence, expression and orthology information for ORs in >50 species (Crasto *et al.*, 2002). The Human Olfactory Data Explorer (HORDE), which is also accessible through ORDB, is specially formulated for human ORs to include practical genomic information (Safran *et al.*, 2003). In addition to the sequence and expression information, many studies have provided functional profiles of ORs using a diverse array of techniques. Recently, large-scale matching of ORs to ligands was made possible through the establishment of

a number of heterologous expression systems. With an increasing number of orphaned mammalian ORs, it became imperative to integrate these data into an all-inclusive database. However, to our knowledge, there is yet no such chemical biology database detailing OR-ligand pairs in mammals.

While extracting OR-ligand pairs from literature to a database may provide accessibility, the power to infer novel candidate receptors for a given ligand would ultimately fuel further research in the field. Several studies used computational models to predict olfactory responses based on existing ones. For example, Khan *et al.* employed principle component analysis (PCA) for the reduction of the dimensionality of percept and of physicochemical descriptors to explain the pleasantness of novel molecules by their properties (Khan *et al.*, 2007). The same group also demonstrated that a multidimensional metric may be a better predictor of differences in olfactory neuronal activities (Haddad *et al.*, 2008). Recently, using an OR heterologous expression system that supports efficient OR functional activation upon ligand stimulation, we robustly de-orphaned 62 human and mouse ORs and used the data to develop a model for predicting odorant-receptor activation (Saito *et al.*, 2009).

In this study, we present for the first time a web server for predicting the ORs for small-molecule compounds, made possible by the permutation of two individual models aimed at defining an odorant and then predicting its candidate ORs. In benchmark studies, the model for odorant determination shows a high accuracy of ~95% on the test data and the model for receptor recognition has improved performance by ~20% in comparison with the previous method (Saito *et al.*, 2009). More importantly, the prediction of the ORs for several odorants using ODORactor was experimentally validated in the study. Thus, the dual prediction capability and the user-friendly interface of this web server can be readily applied to experiments involving mammalian ORs to serve as a starting point for identifying the respective receptors for chemicals of interest by allowing a more informed approach in selecting odorants and receptors.

## 2 SERVER DESCRIPTION

### 2.1 Olfactory resource

All olfactory data collected from scientific literature and resources are stored in a MySQL database and available in the 'Browse' page of the ODORactor server (Supplementary Fig. S1). It contains two types of data: (i) odorants; (ii) ORs in both human and mouse. In total, 3038 odorants and 1608 ORs are curated and fully annotated by the server developers and experts in the field. Of the 1608 ORs

\*To whom correspondence should be addressed.

<sup>†</sup>The authors wish it to be known that, in their opinion, the first three authors and the eleventh author should be regarded as joint First Authors.

in ODORactor, 1451 (90.23%) proteins are included by ORDB and HORDE. Contrary to the receptors, >96.8% of the odorants are not previously covered under the 'Odor Database' in ORDB. Users can access information on 'odorant' annotation, such as molecular weight, formula, SMILES, CAS ID. Furthermore, the known ORs for the odorant are listed with original references (Supplementary Fig. S1). Alternatively, users have the option of retrieving relevant 'receptor' information, including all verified odorants and external links to GenBank and Swissprot.

## 2.2 Odorant Identification

Odorant Identification in ODORactor performs on two individual steps: odorant verification and OR recognition, which are thoroughly described in the 'Materials and Methods' of the Supplementary Material. Four input options are provided for a certain organic molecule query, viz., entering the SMILES under 'SMILES', uploading a molecular file in the MDL mol format under 'Structure File', sketching the structure under 'Structure' and selecting a deposited CAS ID under 'Inclusion'. After defining query molecule, two mandatory parameters must be set in order to submit a job: specifying a 'Job Name' enables the users to easily locate their queries in the Job Queue and selecting the 'Organism' for the retrieval of potential ORs from our OR repertoire against the query molecule. Once the run is submitted, a transition window pops up with an associated Job ID. Each job submission is provided with unique Job ID based on the current date and time that serves as a permanent bookmarkable link to the data. The users can apply the unique Job ID or Job Name to track the progress of the calculation in the 'Job Queue' page of ODORactor. A typical run of ODORactor job takes 15–30 s, depending on the complexity of the input molecule. Therefore, huge molecules with more than 250 atoms are not allowed to run in current version of ODORactor due to extended running time. Upon completion of a job, a button labeled 'Finished' emerges in the 'Job Queue' page and can redirect the users to the result. In the future, we plan to update the list of ORs every 6 months in order to provide the latest receptors to screen.

The output in ODORactor is split into three main sections, namely, 'Details of Molecule', 'Receptor List' and 'Download'. 'Details of Molecule' provides the first result of ODORactor on odorant verification and some important descriptors of the molecule are also shown in the panel. If the query molecule is predicted as an odorant in 'Type', the subsequent prediction of ORs will be run in ODORactor to give the second result of ODORactor on OR recognition, shown in 'Receptor List', which includes all predicted receptors along with their overall confidence scores in probability. The confidence score is provided for each potential receptor–ligand interaction as described in the 'Materials and Methods' of Supplementary Material and ranges from 0% to 100%, with 0% indicating maximum confidence for noninteraction and 100% indicating maximum confidence for interaction between odorant and OR. For example, a potential OR for a query molecule with the estimated probability of 90% should be more likely to be the receptor than one with a probability of 60%. Therefore, only potential ORs of the query molecule with probabilities >50% are ranked as positives in descending order. In addition to the visualizations, all results and annotations in ODORactor can be downloaded as XML text for analyses under the 'Download' panel.

We have extensively tested our algorithm (Supplementary Tables S1 and S2) and two examples of these are provided in the Help page. To further test the reliability of ODORactor server

in practice, an in-house fragment library, which is composed of 210 small organic compounds and used to screen the fragment-based hit discovery, was evaluated for their olfactory responses among the mouse ORs via the server (Supplementary Table S3). Among the compounds, three were predicted to be odorous and their potential receptors were outputted. Furthermore, the interactions between the three compounds and their respective top five receptors identified via the ODORactor were evaluated using a luciferase reporter gene assay (see 'Materials and Methods' of Supplementary Material). Among the predicted candidates, seven OR-odorant pairs were experimentally validated (Supplementary Fig. S2 and Table S4). Therefore, we believe that our computational models represent an effective means of deciphering olfactory coding and could be a useful tool for both basic olfactory research and industry-scale receptor-odorant matching.

## 3 DISCUSSION

We have developed the ODORactor web server that functions as a platform for the advanced identification of chemical odorants and their cognate receptors in combination with an olfactory repository covering annotated odorants, receptors and their interactions. To the best of our knowledge, this web server is the first of its kind and will be of considerable value to scientists interested in olfaction. We have demonstrated that ODORactor is capable of identifying novel odorants and their receptors. For example, we predicted and subsequently validated additional ligands for MOR271-1 and MOR272-1, which were previously matched to ligands of diverse structures. The potencies of these newly identified ligands, as assessed by their respective EC<sub>50</sub> values, fall within the same previously described range for each of the receptors. In addition, using ODORactor, we were able to successfully de-orphan two mouse ORs, MOR244-3 and Olfr42, matching them to 3 and 1 ligands, respectively. Finally, a prominent strength of this web service is its ongoing integration of up-to-date information of both odorant and its receptors from literature for improving its prediction power. The ODORactor server is available at <http://mdl.shsmu.edu.cn/ODORactor>.

## ACKNOWLEDGEMENTS

We thank Dr Hiroaki Matsunami and Dr Joel Mainland for fruitful discussions.

**Funding:** National Natural Science Foundation of China (21002062 and 30970981, in part); Shanghai PuJiang Program (10PJ406800 and 09PJ1406900, in part).

**Conflict of Interest:** none declared.

## REFERENCES

- Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, **65**, 175–187.
- Crasto, C. *et al.* (2002) Olfactory receptor database: a metadata-driven automated population from sources of gene and protein sequences. *Nucleic Acids Res.*, **30**, 354–360.
- Haddad, R. *et al.* (2008) A metric for odorant comparison. *Nat. Methods*, **5**, 425–429.
- Khan, R.M. *et al.* (2007) Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *J. Neurosci.*, **27**, 10015–10023.
- Safran, M. *et al.* (2003) Human gene-centric databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. *Nucleic Acids Res.*, **31**, 142–146.
- Saito, H. *et al.* (2009) Odor coding by a mammalian receptor repertoire. *Sci. Signal*, **2**, ra9.