DeepNose: Using artificial neural networks to represent the space of odorants

Ngoc B. Tran [1] **Daniel R. Kepple** [1] **Sergey A. Shuvaev** [1] **Alexei A. Koulakov** [1]

Abstract

The olfactory system employs an ensemble of odorant receptors (ORs) to sense odorants and to derive olfactory percepts. We trained artificial neural networks to represent the chemical space of odorants and used this representation to predict human olfactory percepts. We hypothesized that ORs may be considered 3D convolutional filters that extract molecular features and, as such, can be trained using machine learning methods. First, we trained a convolutional autoencoder, called DeepNose, to deduce a low-dimensional representation of odorant molecules which were represented by their 3D spatial structure. Next, we tested the ability of DeepNose features in predicting physical properties and odorant percepts based on 3D molecular structure alone. We found that, despite the lack of human expertise, DeepNose features often outperformed molecular descriptors used in computational chemistry in predicting both physical properties and human perceptions. We propose that DeepNose network can extract *de novo* chemical features predictive of various bioactivities and can help understand the factors influencing the composition of ORs ensemble.

1. Introduction

The olfactory system has evolved to sense and recognize a large number of chemicals present in the environment. Olfactory receptors (ORs) act as molecular sensors that bind odorants and convey information about their structure to higher brain regions. ORs are activated by molecules in a combinatorial manner: one receptor can bind to many molecules, and one molecule can interact with many receptors (Buck & Axel, 1991; Malnic et al., 1999; Saito et al., 2009; Soucy et al., 2009). The number of ORs varies amongst species: from 350 in humans to about 1000 in other mammals (Olender et al., 2004). Here, we study the factors influencing the OR ensemble composition.

We trained multilayer (deep) convolutional neural networks (CNNs) to map the space of molecules and used the derived representation to predict human perceptual responses. CNNs are multilayer neural networks that use ensembles of spatial filters of increasing complexity to extract features of the objects present in the input (Krizhevsky & Hinton, 2012). Networks based on CNNs are the best performing models for image recognition, many of which perform better at classifying images than humans (Szegedy et al., 2015; He et al., 2016). The performance of CNNs in 2D image recognition tasks suggests that they are well-suited for learning other types of spatial data, such as 3D molecular structures. Deep networks have indeed been applied to classify molecules, including the predictions of pharmacokinetics properties, toxicity, and protein-ligand interactions (Unterthiner et al., 2016; Xu et al., 2017; Ragoza et al., 2017; Torng & Altman, 2017). CNNs were used to represent the SMILES string of molecules (Gómez-Bombarelli et al., 2018). In this study, we derived CNNs that can recognize 3D molecular structures and predict human olfactory percepts.

Our main hypothesis is that the OR ensemble forms a set of 3D filters that interact with molecules in real space. In this regard, molecules play the role of 3D images (Figure 1) and ORs can be viewed as analogs of receptive fields, such as edge detectors of the visual system (Hubel & Wiesel, 1962). Both the visual receptive fields and the OR ensemble have been shaped by evolution to sense relevant features present in the stimuli. We thus hypothesize that the *in silico* OR ensemble can be optimized using conventional machine learning methods, such as backpropagation (Rumelhart et al., 1985), to accurately represent the space of molecules. Using this approach, machine learning may help us understand the factors influencing OR evolution and construct a realistic model of OR-to-ligand binding.

One of the biggest limitations in building a robust predictive computational model of olfactory processing is the lack of large-scale data on OR binding affinities and perceptual odorant descriptors. Recent efforts have been successful in generating data for hundreds of odorants (Dravnieks, 1985; Saito et al., 2009; Keller et al., 2017), while modern machine learning methods require thousands to millions of data

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. Correspondence to: Alexei Koulakov <koulakov@cshl.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

points to be successful (Lecun et al., 2015). To overcome this problem, we trained a stacked deep autoencoder (Hinton & Salakhutdinov, 2006), which we called DeepNose, to replicate 3D molecular structures (Figure 2). Training the autoencoder relies on unsupervised learning, and, as such, can be implemented using large-scale unlabeled datasets containing 3D molecular structures (Bolton et al., 2011).

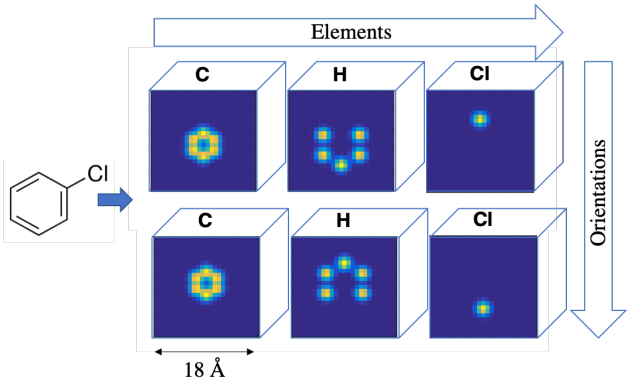

Figure 1. Molecules as 5D inputs. Each atom is represented by a Gaussian cloud that is sampled using a 3D square grid. One 3D channel is dedicated to each element. This 'element' channel constitutes the fourth dimension. Different orientations of the molecule constitute the fifth dimension. In this example (chlorobenzene), the molecule is converted into a grid of size 18 Å, with the resolution of 0.5 Å, three elements, and two orientations. The network input is of size $37^3 \times 3 \times 2$.

We found that the autoencoder can be trained to fully and accurately replicate input molecules. Therefore, DeepNose can be viewed not only as a functional analog of the ORs ensemble, but also as a data-driven and reversible chemical feature extractor. We further tested the biological relevance of the extracted features by predicting molecules' chemical and biological properties (Table 1). The resulting accuracies were comparable to or higher than results obtained using Dragon chemical descriptors, state-of-the-art molecular features widely used in odorant percept prediction (Keller et al., 2017; Todeschini & Consonni, 2010; Shang et al., 2017).

2. Results

Molecules in the environment can be viewed as 3D objects that are recognized by the olfactory system. Similarly to images, each molecule can be represented by separate 'color' channels that contain information about the spatial distribution of six elements: C, H, O, N, S, and Cl. This forth dimension, called here 'CHONSCl', is analogous to RGB channels representing color. In contrast to visual stimuli, odorants are presented to the olfactory system in multiple copies simultaneously, differing in orientations and conformations, which constitute the fifth dimension of our input. We therefore represent each input molecule by a 5D object, including three spatial dimensions, one 'color' CHONSCl, and one orientation dimension (Figure 1).

2.1. DeepNose autoencoder can be trained to accurately represent the space of molecules

To take advantage of the large dataset containing $\approx 10^7$ unlabeled molecular structures (PubChem3D (Bolton et al., 2011), Table 1), and to produce a compressed representation of the space of molecules, we designed a stacked autoencoder (Hinton & Salakhutdinov, 2006; Rumelhart et al., 1985) to accurately represent 3D molecular structures. The autoencoder consists of two convolutional neural networks: an encoder, which converts each molecular structure into a feature vector, and a decoder, which recovers the spatial structure (Figure 2A, Table 2). We restricted the feature vector to have a lower dimensionality than the input, and, therefore, induced the network to learn a latent representation that captures the most statistically salient information in the molecular structures. Because it is not necessary and feasible to present all available molecular shapes to the autoencoder, we restricted the training set to 10^5 molecules randomly selected from the PubChem3D database. We verified that the result does not depend on the selection of molecules. After training using a set of 10^5 molecules and 4×10^4 batches (48 molecules/batch), we achieved a test set Pearson correlation of $r = 0.98$ (Figure 2B). Even though the latent representation used ≈ 45 times fewer neurons than the input, it can capture the majority of the input information and generalize to novel molecules. We also calculated the correlation coefficients specific to each element channel. We observed that C and H channels learn substantially faster than the other channels. This might be due to the higher frequency of C and H atoms in the training set. Nonetheless, autoencoder was able to reconstruct atoms of every element to a correlation of $r > 0.95$ (Figure 2B and C). Overall, our

Table 1. Datasets used for the DeepNose autoencoder and classifier.

Dataset	Purpose	Activity	Number of molecules
PubChem3D	Autoencoder	N/A	$10^5/10^7$
ESOL	Classifier	Water solubility	1144
Good Scents	Classifier	580 semantic olfactory descriptors	3826
Flavornet	Classifier	197 semantic olfactory descriptors	738

autoencoder network was capable of accurately replicating small molecules (< 350 g/mol) in both our training and test sets based on a restricted number of features.

Table 2. Autoencoder network architecture.

	Input	Weights	Output
1	$18^3 \times 6 \times 1$	$5^3 \times 6 \times 12$	$14^3 \times 12 \times 1$
2	$14^3 \times 12 \times 1$	$5^3 \times 12 \times 24$	$10^3 \times 24 \times 1$
3	$10^3 \times 24 \times 1$	$5^3 \times 24 \times 48$	$6^3 \times 48 \times 1$
4	$6^3 \times 48 \times 1$	$5^3 \times 48 \times 96$	$2^3 \times 96 \times 1$
5	$2^3 \times 96 \times 1$	$5^3 \times 96 \times 48$	$6^3 \times 48 \times 1$
6	$6^3 \times 48 \times 1$	$5^3 \times 48 \times 24$	$10^3 \times 24 \times 1$
7	$10^3 \times 24 \times 1$	$5^3 \times 24 \times 12$	$14^3 \times 12 \times 1$
8	$14^3 \times 12 \times 1$	$5^3 \times 12 \times 6$	$18^3 \times 6 \times 1$

2.2. DeepNose features accurately predict water solubility

We next asked whether DeepNose features are relevant to the molecules' physicochemical properties. To this end, we used the ESOL dataset (Table 1) that contains information about water solubility for 1144 molecules to train a neural network classifier (Delaney, 2004). The classifier is composed of the autoencoder's convolutional encoder layers, a consolidation layer that averages responses of autoencoder features over spatial and orientation variables, and three fully connected classifier layers (Figure 3A, Table 3). In constructing our network, we took inspiration from the structure of the olfactory system. Thus, the goal of the consolidation layer is to represent the responses of olfactory sensory neurons. In the mammalian olfactory system, each olfactory sensory neuron produces only one type of OR and conveys information about the odorant to the olfactory networks (Malnic et al., 1999). To implement this constraint in our network, since autoencoder features are purported to represent activities of individual *in silico* OR types, and since responses of olfactory sensory neurons are not sensitive to the presentation angle and position of each molecule, we averaged autoencoder feature activity over ligands' spatial and orientation variables. To represent each molecule as input, we used its 3D structure replicated at 320 orientations that uniformly sampled three solid rotation angles forming a "magic number configuration" (Wales & Ulker, 2006). We found that the magic number representation provides better classification accuracy than the one based on 1000 random orientations. To summarize, the classifier receives each molecules' 5D input structure of dimension $18^3 \times 6 \times 320$, generates a set of molecular features using the encoder layers of DeepNose autoencoder, averages these features over space and orientation using the consolidation layer, and passes the resulting responses to the fully connected classifier layers (Figure 3A, Table 3, $N_{dim} = 1$). We trained the classifier using 10^5 iterations and evaluated the performance of the test set using the network with the highest validation set's correlation. DeepNose features, when paired with a neural networks classifier, are predictive of the water solubility with the accuracy of $r = 0.96$ (Figure 3B and C).

Table 3. Classifier network architecture. Layers with fixed weights are indicated by $*$. L_1 and L_2 are the number of hidden units in the fully connected layers. For the ESOL dataset, $L_1 = 100$ and $L_2 = 50$. For the Good Scents and Flavornet datasets, $L_1 = 200$ and $L_2 = 100$. N_{dim} is the dimension of the network output and has values 1, 10, and 25 for the ESOL, Flavornet, and Good Scents datasets, respectively.

	Input	Weights	Output
$1*$	$18^3 \times 6 \times 320$	$5^3 \times 6 \times 12$	$14^3 \times 12 \times 320$
$2*$	$14^3 \times 12 \times 320$	$5^3 \times 12 \times 24$	$10^3 \times 24 \times 320$
$3*$	$10^3 \times 24 \times 320$	$5^3 \times 24 \times 48$	$6^3 \times 48 \times 320$
$4*$	$6^3 \times 48 \times 320$	$5^3 \times 48 \times 96$	$2^3 \times 96 \times 320$
$5*$	$2^3 \times 96 \times 320$	Consolidation	$96/768 \times 1$
6	$96/768 \times 1$	$96/768 \times L_1$	$L_1 \times 1$
7	$L_1 \times 1$	$L_1 \times L_2$	$L_2 \times 1$
8	$L_2 \times 1$	$L_2 \times N_{dim}$	$N_{dim} \times 1$

2.3. Analysis of olfactory perceptual spaces

It was previously argued that the human olfactory space can be viewed as a continuous, curved, and low-dimensional manifold (Koulakov et al., 2011). We used two olfactory perceptual datasets, Flavornet and Good Scents, to embed the space of olfactory perceptual objects (Acree & Arn; The Good Scents Company). These two datasets contain information about human responses to a large number of molecules that are represented by semantic descriptors (738 molecules/197 descriptors for Flavornet and 3826 molecules/580 descriptors for Good Scents, Table 1). Because these responses are binary and sparse, for the purposes of obtaining a robust predictor, it is more relevant to obtain a low-dimensional approximation of the datasets and to represent each molecule by a dense coordinate along each of the dimensions (Figure 4A, Figure 5A). To reduce the dimensionality of each dataset, we used the Isomap algorithm (Tenenbaum et al., 2000). Semantic descriptors that are significantly enriched in the positive or negative directions of the first three dimensions are illustrated in Figures 4C-E and 5C-E.

The first perceptual dimension with the highest included variance has been previously argued to represent a molecule's pleasantness (Kepple & Koulakov, 2017). Indeed, the first perceptual dimensions of the overlapping set of molecules between Flavornet and Good Scents datasets have the highest correlation coefficient amongst all dimension pairs ($r = 0.54$, data not shown). Interestingly, we also found that the second dimension of Good Scents database is

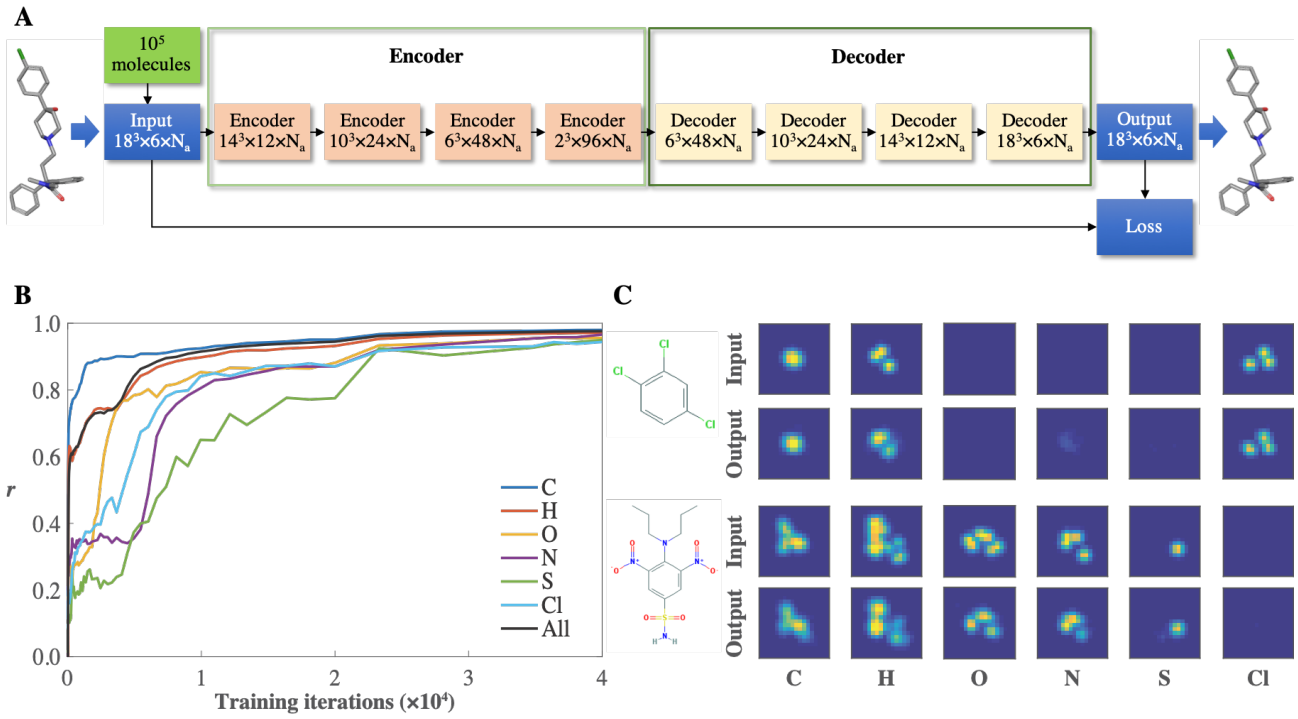

Figure 2. DeepNose autoencoder. (**A**) The autoencoder consists of an encoder and a decoder. The encoder transforms the input into a latent representation, while the decoder performs the inverse operation to reconstruct the spatial structure of input molecule. The autoencoder is trained to minimize the difference between the input and the reconstruction (the number of random orientations used was $N_a = 1$. (**B**) Pearson correlation between input and output for test set molecules as a function of the number of training iterations (batches, 48 molecules/batch, black). Correlations for individual elements are also shown: carbon (blue), hydrogen (red), oxygen (yellow), nitrogen (purple), sulfur (green) and chloride (cyan). (**C**) Examples of two molecules reconstructed by the autoencoder: trichlorobenzene (top) and oryzalin (bottom). Each box is a 2D projection of the 3D grid for individual elements.

correlated with the third dimension of Flavornet ($r = 0.44$). Overall, we reduced the dimensionality of two discrete semantic datasets of molecules describing human olfactory percepts and found that several principal dimensions in both datasets are correlated for the overlapping set of molecules.

We then tested how robust each datasets Isomap dimensions are to the selection of the set of molecules using resampling techniques. To this end, we selected two random subsets from each dataset with a small overlap (100 molecules), performed low dimensional embedding for each subset separately, and compared Isomap dimensions for the common set of molecules. A large correlation between two subsets ($r \approx \pm 1$) would imply that the given Isomap dimension is robust to the selection of molecules. For the Good Scents dataset, we found that the first three dimensions are more consistent compared to other dimensions (Figure 5B), suggesting that these dimensions are robust to the selection of molecules. For Flavornet, we found that only the first dimension is consistent between the two subsets, and that other dimensions appear uncorrelated between subsampled data (Figure 4B). Overall, we found that the first dimension in both datasets is both consistent between them and is robust to the selection of molecules. Good Scents contains more dimensions that are robust to resampling than Flavornet (3 vs. 1), which can be attributed to a higher number of molecules included in Good Scents compared to Flavornet (3826 vs. 738).

2.4. DeepNose features predict perceptual dimensions with accuracy comparable or higher than Dragon descriptors

We then used our classifier network to predict human percepts (Isomap dimensions) based on 3D molecular shapes. For each molecule in the Flavornet and Good Scents datasets, we used the trained encoder layers and the consolidation layer to produce a feature vector, which were then used as inputs to a fully connected feedforward neural networks (Figure 3A, Table 3, $N_{dim} = 10$ for Flavornet, $N_{dim} = 25$ for Good Scents). To benchmark DeepNose features, we used Dragon chemical descriptors often used in olfactory predictions (Keller et al., 2017; Todeschini & Consonni, 2010; Shang et al., 2017), for comparison. Dragon features include a diverse set of 4885 molecular descriptors

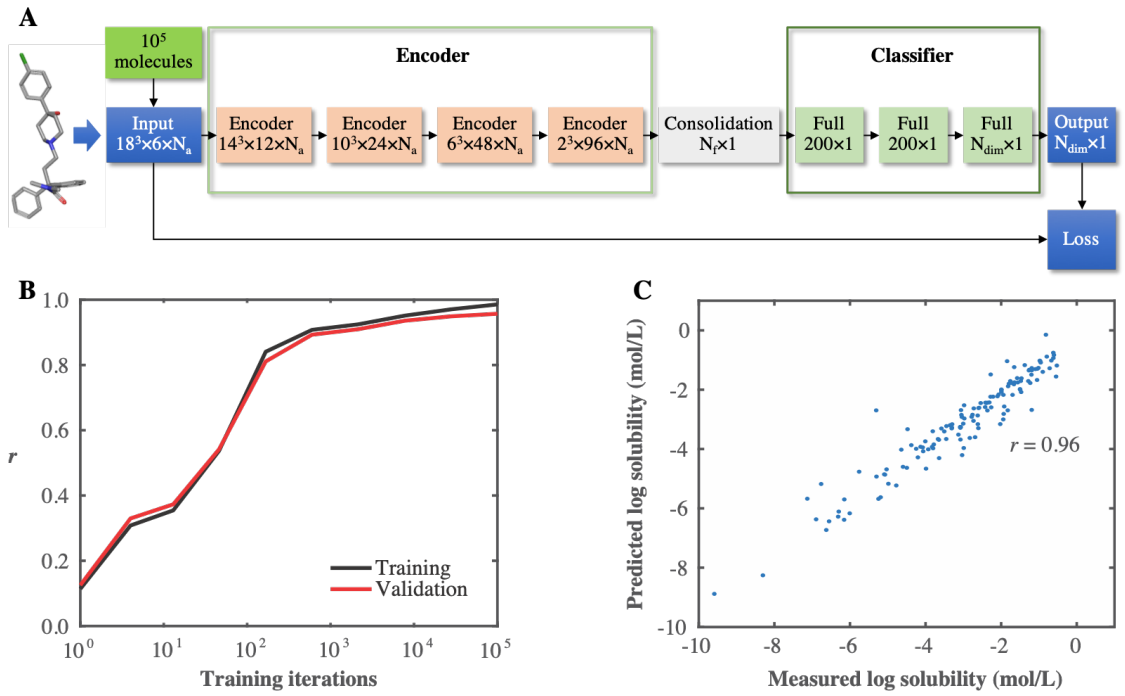

Figure 3. DeepNose classifier and its performance in predicting molecules' water solubility. (**A**) The classifier consists of four encoder layers, one consolidation layer and three fully connected layers. The encoder layers are transferred from the trained autoencoders first four layers; the consolidation layer averages DeepNose features over $N_a = 320$ orientations; and the fully connected layers transform the features into the bioactivity of N_{dim} dimensions. The N_{dim} values of the water solubility (ESOL), Good Scents, and Flavornet datasets are 1, 25, and 10, respectively. (**B**) Average Pearson correlation coefficient of training (black) and validation (red) set molecules for the ESOL dataset as functions of the number of training iterations (batches, 8 molecules/batch). (**C**) The predicted versus measured log water solubility for the test set molecules after training (10^5 iterations, $r = 0.96$).

ranging from simple, such as molecular weight and atomic composition, to more complex ones, such as topological molecular indices (Todeschini & Consonni, 2010). These properties summarize the variables developed in computational chemistry to characterize molecules. We included approximately 2000 Dragon descriptors that had non-trivial values (non-zero variance) for the molecules used. Dragon features were computed for each molecule and used as inputs to the fully connected layers in Figure 3A and Table 3. We found that DeepNose features, when paired with classifier networks, gave predictions of accuracy comparable to or exceeding Dragon descriptors. For example, for the first dimension in Good Scents database (Figure 5C), the median correlation between predicted and observed values is between $r = 0.67 - 0.68$ if DeepNose features are used versus $r = 0.66$ for Dragon features ($p < 0.05$, Figure 5C). For the first dimension of Flavornet, the correlations are between $r = 0.57 - 0.59$ for DeepNose features compared to $r = 0.54$ for Dragon ($p < 0.05$, DeepNose 96 features, Figure 4C). Overall, we suggest that DeepNose autoencoder is able to automatically discover useful chemical features from 3D shapes of molecules alone. These features are useful in predicting human olfactory percepts.

3. Discussion

In this study, we applied neural networks to directly encode information about 3D molecular conformations. Our main hypothesis is that ORs act as 3D filters that are activated by molecules' spatial features. As such, these features can be extracted and ORs 'trained' using machine learning techniques, such as backpropagation. First, we used autoencoder networks to capture the latent space sampled by molecules. The autoencoder, in our implementation, is a convolutional feedforward multilayer neural network which receives 3D molecular conformations as inputs and is trained to faithfully replicate these conformations as outputs. The information about molecular shapes passes through the middle latent layer, which contains about 45 times fewer variables than both inputs and outputs. Our DeepNose autoencoder therefore accomplished its goal using a substantial compression of information. Our assumption is that the ensemble of ORs is under evolutionary pressure to provide a compressed and accurate representation of molecular shapes that is further interpreted by olfactory networks. Under this assumption, we can associate the latent variables produced by our autoencoder with the responses of the ensemble of ORs. We

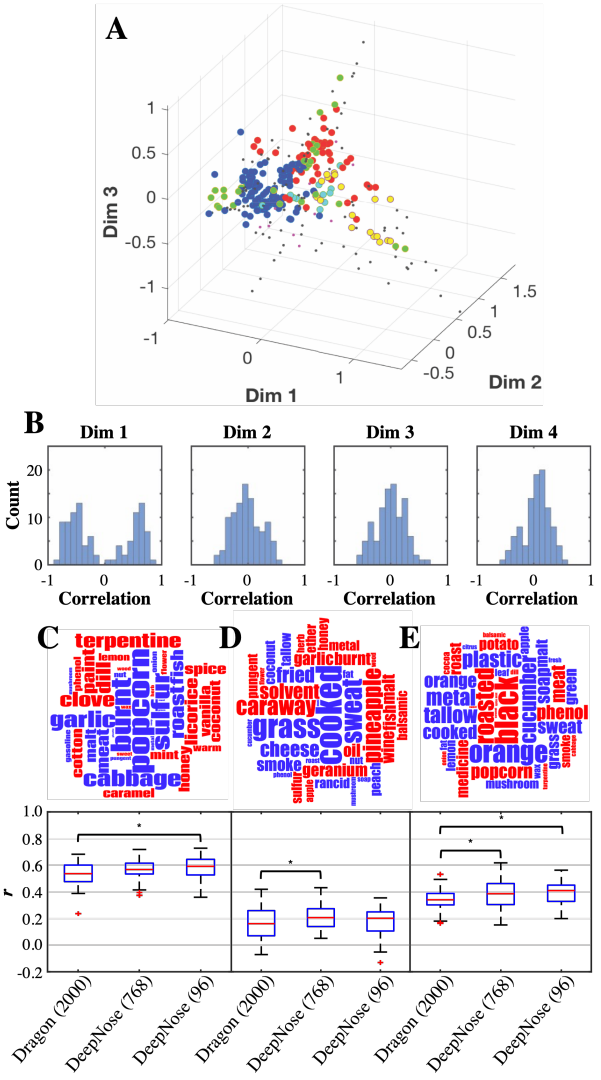

Figure 4. Perceptual dimensions in the Flavornet dataset. (**A**) 3D Isomap embedding of the dataset. Individual molecules (points, circles) are colored according to functional clusters (Oyibo et al., 2018). (**B**) Correlation coefficients of the Isomap dimensions obtained from two halves of the dataset for 100 resamplings show that only the first dimension is robust to the selection of molecules. (**C**-**E**) Top: Word clouds showing semantic descriptors significantly enriched in the positive (blue) or negative (red) direction of the first, second and third Isomap dimensions. Bottom: Prediction accuracy for these dimensions using DeepNose features or Dragon chemical descriptors ($*$: $p < 0.05$, right-tailed Wilcoxon rank sum test).

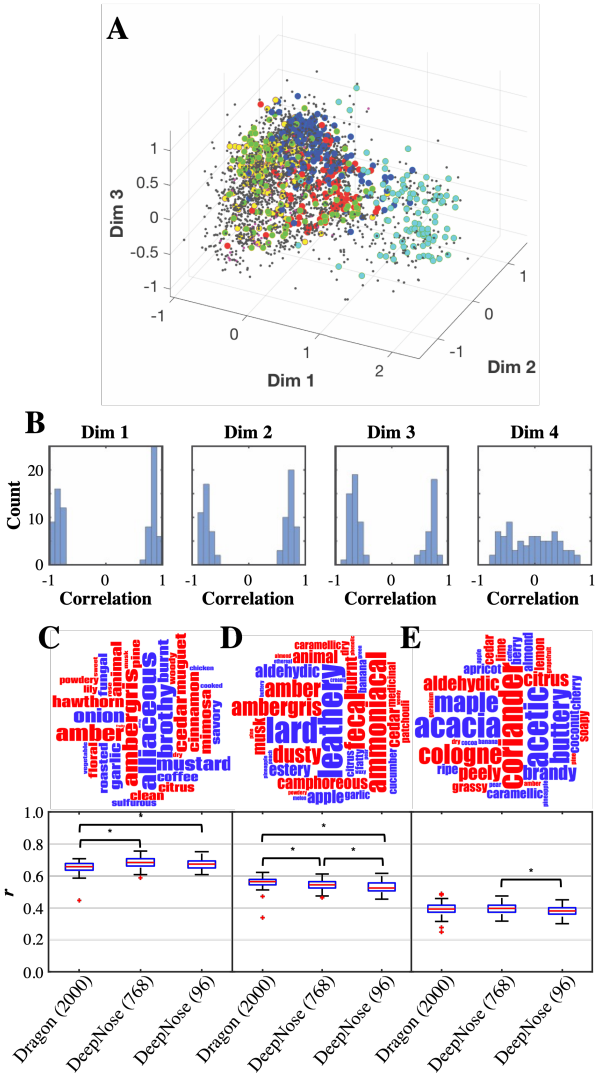

Figure 5. Perceptual dimensions in the Good Scents dataset. (**A**) 3D Isomap embedding of the dataset. Individual molecules (points, circles) are colored according to functional clusters (Oyibo et al., 2018). (**B**) Correlation coefficients of the Isomap dimensions obtained from two halves of the dataset for 100 resamplings show that the first three dimensions are robust to the selection of molecules. (**C**-**E**) Top: Word clouds showing semantic descriptors significantly enriched in the positive (blue) or negative (red) direction of the first, second and third Isomap dimensions. Prediction accuracy for these dimensions obtained using DeepNose features or Dragon chemical descriptors ($*$: $p < 0.05$, right-tailed Wilcoxon rank sum test).

thus propose a framework within which the evolution of ORs can be viewed as neural network training.

Although DeepNose latent features can be interpreted as OR responses, it is much harder to give a biological interpretation to individual layers in our network. We could argue, for example, that the convolutional filters in the first layer of the autoencoder represent individual amino acids within the OR binding pocket, while subsequent layers correspond to assemblies of amino acids of increasing complexity. To make our networks more similar to the biological system, in our classifier network (Figure 3A), we placed a special layer called 'consolidation'. The role of this layer is to compute the output of the encoder in a position and orientation independent manner by averaging latent variables over these

degrees of freedom. This layer is therefore intended to replicate the outputs of olfactory sensory neurons which are not sensitive to the ligand's position and orientation in space.

In our study, we proposed a new way of representing small molecules that is inspired by the interaction between odorants and ORs. At the core of our approach is the realization that each molecule is not a single 3D object but a large number of such objects differing in orientations. Although the visual system is exposed to a single image at a time, olfactory system simultaneously deals with thousands of such 3D images, approaching ORs at different orientations. This is the key difference between the visual and olfactory systems, which is reflected explicitly in our network architecture. Thus, we represented each molecule by a 5D array: 3 spatial, 1 element, and 1 orientation dimensions. Overall, in our study, we adjusted network architecture to reflect the specifics of olfactory stimuli, such as multiple inputs presented simultaneously at different orientations.

We further used the features generated by our network to replicate human olfactory percepts. Because of the need for a large number of molecules in the training set, we used large olfactory perceptual datasets (Good Scents and Flavornet) that place molecules into discrete semantic space and contain thousands of molecules. We described olfactory percepts by the salient dimensions captured by Isomap (Figures 4 and 5A). We added several fully connected classifier layers that crudely approximated olfactory networks, which, in reality, have recurrent structures. We found that the autoencoders 96 latent features yield a good predictive value to the first Isomap perceptual dimension (median $r \approx 0.67 - 0.68$ for the largest olfactory dataset). These values of correlation were comparable to the ones generated by state-of-the-art molecular descriptors used computational chemistry (Dragon, median $r = 0.66$). This finding suggests that DeepNose features represent variables relevant to human olfaction. DeepNose features are computed for each molecule *de novo*, based on its 3D shape only. As such, they can be interpreted as outputs of OR ensemble. We therefore suggest that our networks provide a useful set of molecular descriptors that can also be interpreted as outputs of biophysical ORs. When information about responses of real ORs becomes available for a substantial number of ligands, DeepNose features can be adapted to resemble their biological counterparts.

4. Methods

4.1. Datasets

PubChem PubChem is a publicly available dataset that contains more than 200 million unique molecules (Kim et al., 2016). PubChem3D contains precomputed 3D conformations of molecules. PubChem3D included up to ten conformations per molecule while attempting to maintain coverage and match experimentally-determined 3D structures (Bolton et al., 2011). To consider molecules that are odorant-like, we further filtered for those that weigh less than 350 g/mol and contain up to six elements - C, H, O, N, S, and Cl, or "CHONSCl". We randomly selected 10^5 of these eligible molecules to serve as the training set of the DeepNose autoencoder.

ESOL The ESOL dataset contains 1144 molecules with measured water solubility (Delaney, 2004). We used this dataset to test the ability of DeepNose features to predict water solubility.

Flavornet The Flavornet dataset contains 738 odorant molecules associated with one or more perceptual descriptors, such as flower, lemon, fish, etc. There are 197 of these descriptors (Acree & Arn). We used this dataset to test the ability of DeepNose features to predict odorant percepts.

Good Scents The Good Scents dataset contains 3826 perfumes associated with one or more semantic perceptual descriptors, such as fruity, woody, citrus, etc. There are 580 of these descriptors (The Good Scents Company). We used this dataset to test the ability of DeepNose features to predict odorant percepts.

4.2. Dimensionality reduction of perceptual datasets

We used Isomap to reduce dimensionality of perceptual descriptors in Flavornet and Good Scents datasets (Tenenbaum et al., 2000; Kepple & Koulakov, 2017). For each dataset, we constructed a graph in which each molecule is a node and two molecules are connected by an edge if they share a semantic descriptor. The distance between two connected molecules is calculated based on the overlap between those two molecules percepts, according to the following equations: $E_{ij} = 1 - O_{ij}/\sqrt{O_{ii} \times O_{jj}}$. Here O_{ij} is the number of semantic descriptors shared by molecules i and j. We then used Isomap to compute geodesic distance between all pairs of molecules and embed them into a low dimensional space. We tested the ability of DeepNose features to predict these perceptual dimensions against molecular descriptors produced by the Dragon 6 software (Todeschini & Consonni, 2010).

One way to visualize what is represented by each Isomap dimension is to identify perceptual descriptors that are enriched in the positive or negative directions of particular dimensions. For example, if a descriptor is associated with N molecules, that descriptor is mapped to N values on the first Isomap dimension. We compared the distribution of these N Isomap values associated with each semantic descriptor to the distribution of all descriptors using the non-parametric Mann-Whitney U test. We obtained a set

of semantic descriptors that are significantly enriched in the positive or negative directions of each Isomap dimension. These descriptors are displayed as word clouds (Figure 4C-E, Figure 5C-E).

Second, we tested how robust Isomap dimensions are to subsampling within each dataset. To do that, we selected two random subsets from each dataset with a small overlap (100 molecules), performed low dimensional embedding for each subset separately, and compared Isomap dimensions for the common set of molecules. We repeated this analysis for 100 times and calculated the Pearson correlation coefficients between the two subsets for each dimension. For each dimension, if correlations are near 1 or -1, this Isomap dimension may be considered robust to molecular composition. The distributions of these correlations are shown in Figures 4B and 5B.

Since 520 molecules are present in both Flavornet and the Good Scents datasets, we also tested how much these two datasets correlate for these molecules. We calculated the Pearson correlation coefficient between Isomap dimensions in two datasets for overlapping set of molecules. We found that Isomap dimensions 1 and 2 of Good Scents is significantly correlated ($r = 0.4 - 0.5$) with dimensions 1 and 3 of Flavornet, respectively.

4.3. Input representation

Just as images are represented as 2D objects grouped into separate RGB channels, in our approach, molecules are presented to the network as 3D objects grouped into six channels corresponding to chemical elements (C, H, O, N, S, Cl, or CHONSCl). Spatial distribution of atoms of the same type is represented by a 3D grid. We used grids that are 17Å on each side, with a resolution of 1 Å. Each atom is a 3D Gaussian distribution centering around its xyz-coordinates provided by PubChem3D (Kim et al., 2016). We use a fixed standard deviation of 1 Å for all elements. To incorporate different orientations of the molecules, we sample N_a rotation angle for every training instance of the molecules, where $N_a = 1$ for the autoencoder network and $N_a = 320$ for the classifier network.

4.4. Training DeepNose autoencoder

The goal of autoencoder is to produce a compact representation of the space of molecules. It consists of two convolutional neural networks: an encoder, which converts each molecular structure into a feature vector, and a decoder, which recovers the structure (Figure 2, Table 2). We restricted the feature vector to have a lower dimensionality than the input, and therefore induced the network to learn a latent representation that captures the most statistically salient information in the molecule's structures. We refer to this latent representation as DeepNose features.

One training instance can include one or more orientations, and the same weights are used for each orientation. The autoencoder is trained to minimize the l_2 norm error using the backpropagation algorithm (Rumelhart et al., 1985). We used a learning rate of 2×10^{-7} and a momentum of 0.9. We trained the network using 4×10^4 batches, with each batch containing 48 molecules. We used a training set of randomly selected 10^5 PubChem3D molecules and cross-validated our network with a test set of randomly selected 92 molecules in the ESOL dataset. We evaluated the network performance by calculating the pixel-to-pixel error and correlation between the original molecules and the reconstructed molecules in the test set.

4.5. DeepNose classifier

The input of the classifier is a 5D object, similar to that of the autoencoder. We pair the trained encoder layers of our DeepNose autoencoder with fully connected feedforward layers to predict bioactivities of molecules (Table 1), including solubility and perceptual coordinates. Since molecules adopt multiple conformations and orientations, we made a simplifying assumption that the different orientations contribute with equal probability to its activity. We therefore added a consolidation layer, which averages network activity over multiple orientations of the same molecules within the same latent feature. We used the "magic number configuration" that uniformly sampled three solid rotation angles to get $N_a = 320$ orientations and found that this representation provide better classification accuracy than using $N_a = 1000$ random orientations (Wales & Ulker, 2006).

Thus, each molecule is converted to a vector of either 96 DeepNose features (after consolidating over space and orientation), 768 DeepNose features (after consolidating over orientation), or 2000 non-trivial Dragon features (non-zero variance). These features are fed into the fully connected classifier layers. The classifier (Figure 3A, Table 3) is trained to minimize the l_2 norm error. Weights in the the first four layers are transferred from the autoencoder's encoder layers and fixed. For the ESOL dataset, we used an l_1 regularizer $\lambda = 10^{-3}$ and a learning rate of 10^{-3}. For the Flavornet and Good Scents datasets, we used an l_2 regularizer $\lambda = 10^{-3}$ and a learning rate of 10^{-3}. For all datasets, we trained the networks for 10^5 batches, with each batch containing 8 molecules. We split the dataset into three sets, 70%, 15% and 15%, for training, validation and test sets, respectively. We evaluated the test set using the models with the highest validation set's correlation.

References

Acree, T. and Arn, H. Flavornet and human odor space. URL http://www.flavornet.org.

Bolton, E. E., Kim, S., and Bryant, S. H. PubChem3D : Conformer generation. *Journal of Cheminformatics*, 3(1): 4, 2011. ISSN 1758-2946. doi: 10.1186/1758-2946-3-4. URL http://www.jcheminf.com/content/3/1/4.

Buck, L. and Axel, R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell*, 65(1):175–187, 1991. ISSN 00928674. doi: 10.1016/0092-8674(91)90418-X.

Delaney, J. S. ESOL: Estimating Aqueous Solubility Directly from Molecular Structure. *Journal of Chemical Information and Modeling*, 44(3):1000–1005, 2004.

Dravnieks, A. *Atlas of odor character profiles*. ASTM, Philadelphia, PA, 1985. ISBN 0803104561.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, oct 2018. ISSN 23747951. doi: 10.1021/acscentsci.7b00572. URL http://arxiv.org/abs/1610.02415.

He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.90.

Hinton, G. E. and Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(July):504–508, 2006. ISSN 0036-8075. doi: 10.1126/science.1127647.

Hubel, D. H. and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., Mainland, J. D., Ihara, Y., Yu, C. W., Wolfinger, R., Vens, C., Schietgat, L., De Grave, K., Norel, R., Stolovitzky, G., Cecchi, G., Vosshall, L. B., and Meyer, P. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 2017. doi: http://dx.doi.org/10.1101/082495.

Kepple, D. and Koulakov, A. Constructing an olfactory perceptual space and predicting percepts from molecular structure. *arXiv*, pp. 1–8, 2017. URL https://arxiv.org/abs/1708.05774.

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., and Bryant, S. H. PubChem substance and compound databases. *Nucleic Acids Research*, 44(D1):D1202–D1213, 2016. ISSN 13624962. doi: 10.1093/nar/gkv951.

Koulakov, A. A., Kolterman, B. E., Enikolopov, A. G., and Rinberg, D. In search of the structure of human olfactory space. *Frontiers in systems neuroscience*, 5(September):65, 2011. ISSN 1662-5137. doi: 10.3389/fnsys.2011.00065. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3173711{&}tool=pmcentrez{&}rendertype=abstract.

Krizhevsky, A. and Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, pp. 1–9, 2012.

Lecun, Y., Bengio, Y., and Hinton, G. E. Deep learning. *Nature*, 521:436–444, 2015. doi: 10.1038/nature14539.

Malnic, B., Hirono, J., Sato, T., and Buck, L. B. Combinatorial receptor codes for odors. *Cell*, 96(5):713–723, 1999. ISSN 00928674. doi: 10.1016/S0092-8674(00)80581-4.

Olender, T., Feldmesser, E., Atarot, T., Eisenstein, M., and Lancet, D. The olfactory receptor universe - From whole genome analysis to structure and evolution. *Genetics and Molecular Research*, 2004. ISSN 16765680. doi: 0003[pii].

Oyibo, H., Cao, C., Ferrante, D. D., Zhan, H., Koulakov, A., Enquist, L., Dubnau, J., and Zador, A. A computational framework for converting high-throughput DNA sequencing data into neural circuit connectivity. *bioRxiv*, pp. 1–13, 2018.

Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J., and Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. *Journal of Chemical Information and Modeling*, 57(4):942–957, 2017. ISSN 15205142. doi: 10.1021/acs.jcim.6b00740.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. (V), 1985.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H., and Mainland, J. D. Odor Coding by a Mammalian Receptor Repertoire. *Science Signaling*, 2(60):ra9–ra9, 2009. ISSN 1937-9145. doi: 10.1126/scisignal.2000016. URL http://stke.sciencemag.org/cgi/doi/10.1126/scisignal.2000016.

Shang, L., Liu, C., Tomiura, Y., and Hayashi, K. Machine-Learning-Based Olfactometer: Prediction of Odor Perception from Physicochemical Features of Odorant Molecules. *Analytical Chemistry*, 89(22): 11999–12005, 2017. doi: 10.1021/acs.analchem.

7b02389. URL https://doi.org/10.1021/acs.analchem.7b02389.

Soucy, E. R., Albeanu, D. F., Fantana, A. L., Murthy, V. N., and Meister, M. Precision and diversity in an odor map on the olfactory bulb. *Nature Neuroscience*, 12(2):210–220, 2009. ISSN 10976256. doi: 10.1038/nn.2262.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015. ISBN 9781467369640. doi: 10.1109/CVPR.2015.7298594.

Tenenbaum, J. B., Silva, V. D., and Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(December):2319–2323, 2000. ISSN 00368075. doi: 10.1126/science.290.5500.2319.

The Good Scents Company. The Good Scents Company Information System. URL http://www.thegoodscentscompany.com.

Todeschini, R. and Consonni, V. *Molecular Descriptors for Chemoinformatics*. 2010. ISBN 9783527628766. doi: 10.1002/9783527628766.

Torng, W. and Altman, R. B. 3D deep convolutional neural networks for amino acid environment similarity analysis. *BMC Bioinformatics*, 18(1):1–23, 2017. ISSN 14712105. doi: 10.1186/s12859-017-1702-0.

Unterthiner, T., Mayr, A., Klambauer, G., Hochreiter, S., Unterthiner, T., Hochreiter, S., Mayr, A., Klambauer, G., Hochreiter, S., Unterthiner, T., and Hochreiter, S. DeepTox : Toxicity Prediction using Deep Learning. *Frontiers in Environmental Science*, 3(February):80, 2016. ISSN 2296-665X. doi: 10.3389/fenvs.2015.00080. URL http://journal.frontiersin.org/Article/10.3389/fenvs.2015.00080/abstract.

Wales, D. J. and Ulker, S. Structure and dynamics of spherical crystals characterized for the Thomson problem. *Physical Review B - Condensed Matter and Materials Physics*, 2006. ISSN 10980121. doi: 10.1103/PhysRevB.74.212101.

Xu, Y., Pei, J., and Lai, L. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *Journal of Chemical Information and Modeling*, 57(11):2672–2685, 2017. ISSN 15205142. doi: 10.1021/acs.jcim.7b00244.