



Transport features predict if a molecule is odorous

Emily J. Mayhew^{a,1} , Charles J. Arayata^b , Richard C. Gerkin^c , Brian K. Lee^d, Jonathan M. Magill^b, Lindsey L. Snyder^b, Kelsie A. Little^b, Chung Wen Yu^b, and Joel D. Mainland^{b,e}

Edited by Aravinthan Samuel, Harvard University, Cambridge, MA; received September 10, 2021; accepted February 11, 2022 by Editorial Board Member John R. Carlson

In studies of vision and audition, stimuli can be chosen to span the visible or audible spectrum; in olfaction, the axes and boundaries defining the analogous odorous space are unknown. As a result, the population of olfactory space is likewise unknown, and anecdotal estimates of 10,000 odorants have endured. The journey a molecule must take to reach olfactory receptors (ORs) and produce an odor percept suggests some chemical criteria for odorants: a molecule must 1) be volatile enough to enter the air phase, 2) be nonvolatile and hydrophilic enough to sorb into the mucous layer coating the olfactory epithelium, 3) be hydrophobic enough to enter an OR binding pocket, and 4) activate at least one OR. Here, we develop a simple and interpretable quantitative model that reliably predicts whether a molecule is odorous or odorless based solely on the first three criteria. Applying our model to a database of all possible small organic molecules, we estimate that at least 40 billion possible compounds are odorous, six orders of magnitude larger than current estimates of 10,000. With this model in hand, we can define the boundaries of olfactory space in terms of molecular volatility and hydrophobicity, enabling representative sampling of olfactory stimulus space.

olfaction | odor space | physical transport | machine learning

The number of molecules humans can smell is disputed, with published estimates ranging from 10,000 (1) to infinitely many (2). Chemical space is vast, and we cannot resolve this dispute until we define the subset of chemical space that has an odor. Without defining the boundaries of this space, we cannot know whether previous research has adequately sampled from it, nor understand how the brain represents it, nor conduct a rational search for novel odorants within it.

To act as an odorant, a molecule must complete a transport process to reach olfactory receptors (ORs) and activate one or more ORs (Fig. 1*A*). Viewing the journey of an odorous molecule to an OR as a mass transport problem (3) points to the types of molecular criteria we must consider (e.g., vapor pressure and hydrophobicity), but because no previous studies have both proposed and empirically tested a biological theory of the chemical criteria for an odorant, the relative importance, precise limits, and interactions of these constraints remain undefined.

Results and Discussion

Building an Odor Classification Model. We set out to learn which chemical criteria separate odorous from odorless molecules by building the simplest accurate model from a diverse pool of molecules and chemical features. A model using only three features that dictate transport capability (boiling point [BP], vapor pressure, and octanol/water partition coefficient log *P*) reliably classifies molecules as odorous or odorless (Fig. 1*B* and *C*). If odor classification can be explained by so few molecular properties, why were the classification rules not previously known? The likely reason is that available data are both noisy and poorly curated; we needed to gather odor classification data from multiple sources and correct errors in the data—in both transport features and odorous/odorless labels—before the three-parameter transport model matched the performance of more complicated models.

The theory that transport properties define odor space was previously proposed almost 40 y ago by Boelens (4), but the manuscript provided extremely limited empirical evidence for the claim. More recent efforts to predict odorous/odorless classification of molecules have relied on existing flavor and fragrance databases (5–7); the biased scope of molecules and classification errors in these databases constrain the success of these models. A recent neural network-trained odorous/odorless classification model advertised high accuracy, but we found that its classification accuracy was poor when tested on our cleaned dataset (area under the receiver operating characteristic curve [AUROC] = 0.626; *SI Appendix, Fig. S1*) (6).

Significance

What makes a molecule have a smell? This simple question represents a significant gap in our understanding of olfaction. To answer it, we trained models to predict whether molecules were odorous based on molecular characteristics and noted which characteristics were needed to make correct predictions. We found that molecules with sufficient volatility and hydrophobicity are generally odorous, which suggests that reaching olfactory receptors is the dominant barrier for prospective olfactory stimuli. Based on these criteria, there are billions of molecules that are likely odorous but have never been smelled. We can now recognize an odorous molecule from its structure, and this guides us to discover new classes of odorants and include all types of odorants in our study of smell.

Author affiliations: ^aDepartment of Food Science and Human Nutrition, Michigan State University, East Lansing, MI 48824; ^bMonell Chemical Senses Center, Philadelphia, PA 19104; ^cSchool of Life Sciences, Arizona State University, Tempe, AZ 85287; ^dBrain Team, Google Research, Cambridge, MA 02142; and ^eDepartment of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104

Author contributions: J.D.M. designed research; E.J.M., C.J.A., J.M.M., L.L.S., K.A.L., and C.W.Y. performed research; E.J.M., C.J.A., R.C.G., B.K.L., C.W.Y., and J.D.M. analyzed data; R.C.G., B.K.L., and J.D.M. revised the manuscript; and E.J.M. wrote the paper.

Competing interest statement: J.D.M. received research funding from Ajinomoto Co., Inc.

This article is a PNAS Direct Submission. A.S. is a guest editor invited by the Editorial Board.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: mayhewem@msu.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116576119/-DCSupplemental>.

Published April 4, 2022.

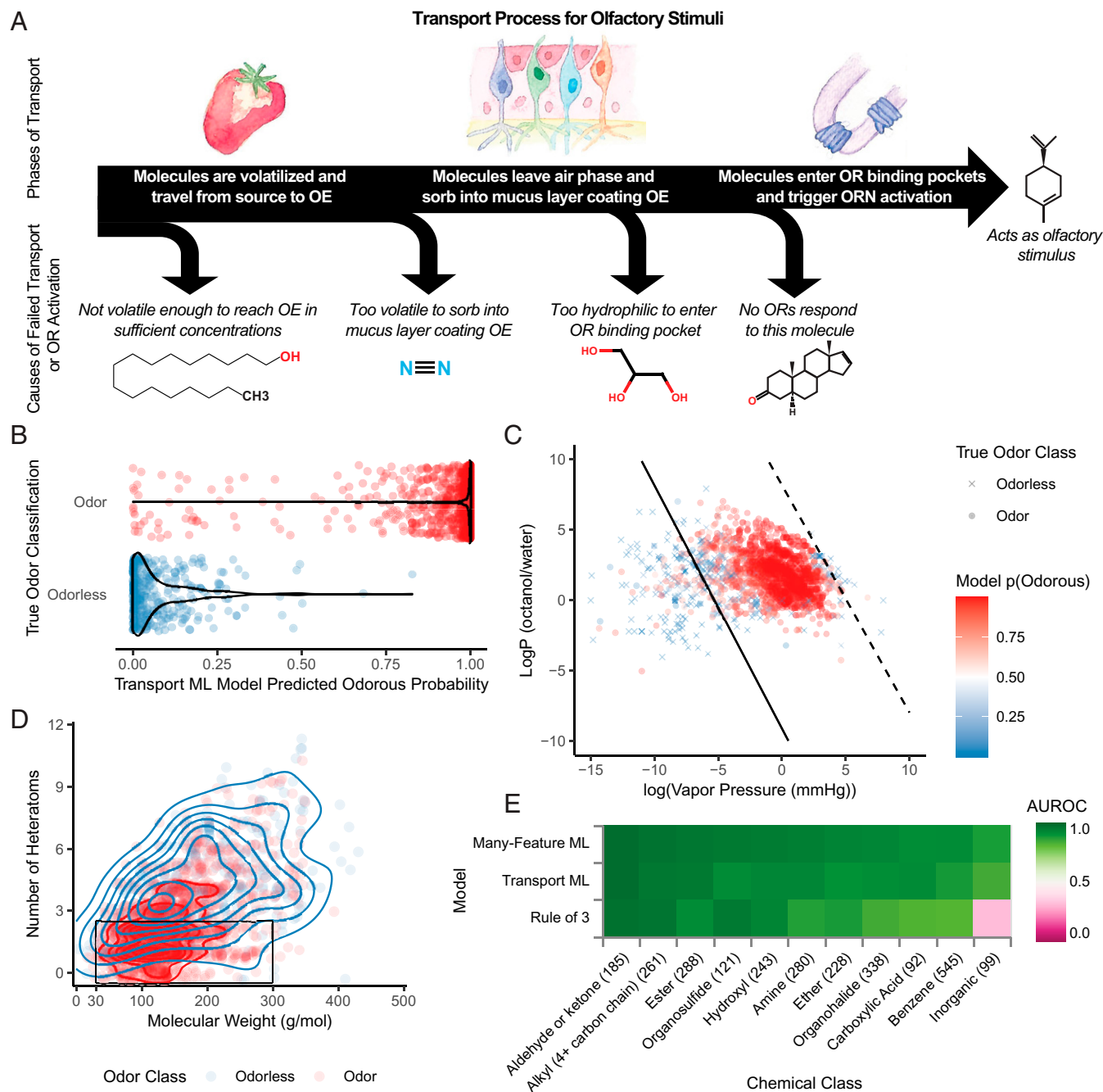


Fig. 1. A model can accurately classify molecules as odorous or odorless based only on transport features. (A) Schematic of the transport process that molecules must complete to act as olfactory stimuli. To elicit an odor, molecules must reach the olfactory epithelium (OE), adsorb into the olfactory mucosa, enter OR binding pockets, and trigger OR neuron (ORN) activation. (B) Transport-feature ML model-generated odorous probabilities for all molecules in the dataset. Each dot represents one molecule colored by the ground truth, and the width of the violin plot is the density of molecules at a given prediction value. (C) Odorous and odorless molecules in transport space. LogP and log(vapor pressure [mmHg]) are plotted for each molecule in the dataset; odorous molecules are represented by circles, and odorless are represented by crosses; molecules are colored by transport-feature ML model-generated odorous probabilities. An LR-generated 50% odorous probability boundary for solids or liquids (Eq. 1) is plotted as a solid line, and the boundary for gases (Eq. 2) is plotted as a dashed line; increasing the value of any feature by X increases the log odds of odorousness by kX, where k is the corresponding model coefficient. (D) Density of odorous and odorless molecules in transport space defined by molecular weight and number of heteroatoms. Each successive contour line indicates a step increase in density (odorous, red = 0.05%; odorless, blue = 0.01%). Each molecule has an integer number of heteroatoms, but these values are jittered along the y axis to better show density. Plotted within the black box, molecules that obey the rule of three are generally odorous. (E) Heat map of mean AUROC generated by the transport ML, many-feature ML, and the rule of three models for molecules of common chemical classes (number of matching molecules in parentheses).

Ultimately, we generated a large and chemically diverse dataset of over 1,900 molecules, classified as odorous (84%) or odorless (16%) through either literature- and web-scraping (literature-classified, 1,796 molecules) or human discrimination tasks and chemical analysis (lab-classified, 128 molecules). We

drew 128 chemically diverse molecules from a pool of 6,000 molecules with safety data. We used *k*-means clustering to group molecules by chemical similarity, then selected at least 4 molecules from each of *k* = 30 clusters to test in the laboratory. Unlike previous efforts, we set aside in advance a test set

of 30 molecules with high-confidence classifications (10 odorless and 20 odorous; classified by human subjects and confirmed through chemical analysis) to measure final model performance. The remaining molecules formed our training set. We applied several common machine learning (ML) algorithms, including logistic regression (LR), support vector machine, random forest, gradient boosting, and extreme gradient boosting (XGB), to train odor classification models. Our models represent molecular structure as a vector of physicochemical features (Dragon v6, Talet; EPI Suite, US Environmental Protection Agency [EPA]) and calculate a probability that the molecule is odorous; all code used to generate models and figures is publicly available (<https://github.com/emayhew/OlfactorySpace>).

We optimized ML models and measured performance using the AUROC, a metric for which 1.0 represents perfect classification and 0.5 represents chance-level classification accuracy. We achieved near-perfect AUROC values in cross-validation (CV) with several algorithms when paired with a synthetic minority oversampling technique to address the imbalance in odorous:odorless training examples; the XGB algorithm produced the highest CV AUROC and is used to generate most of the models and figures in this manuscript. When applied to the held-out test set, a simple three-parameter ML model separates odorous from odorless molecules with no mistakes (AUROC = 1.0; see *SI Appendix*, Fig. S2, for test set performance across algorithms). Retraining transport-feature models and testing on 25 random draws of 30 laboratory-tested molecules yields an AUROC of 0.975 ± 0.028 (mean \pm SD; *SI Appendix*, Table S1). Our laboratory-tested dataset was designed to maximize chemical diversity within safety and availability constraints, so strong predictive performance on these molecules suggests that the model will generalize well to new molecules.

To visualize the odorous region of transport space, we plotted the ML (XGB) transport model predictions for molecules in Fig. 1C. In this plot, we also draw two odorous/odorless boundaries (derived from the simpler but nearly as high-performance LR model): a low-vapor pressure boundary that applies to room-temperature liquids and solids,

$$\text{LogP} > (-1.72)(\log[VP]) - 9.10, \quad [1]$$

where LogP is the log octanol/water partition coefficient and VP is the vapor pressure in mmHg, and a high-vapor pressure boundary that applies to room-temperature gases,

$$\text{LogP} < (-1.61)(\log[VP]) + 8.17. \quad [2]$$

Because the set of laboratory-tested molecules does not include gases, Eq. 1 is sufficient for generating predictions on the test set and yields an AUROC of 0.98; applying both boundaries to the full dataset has somewhat lower accuracy (AUROC = 0.88).

We next compared the performance of our transport ML model to a many-feature ML model trained with over 3,700 physicochemical features. Regularization penalizes use of additional, unnecessary features, and we included high ($\alpha = 1$, $\lambda = 1$) and low weights ($\alpha = 1\text{E-}5$, $\lambda = 1\text{E-}5$) for L1 and L2 regularization in our model tuning parameter set. We found that additional features do not improve test set classification accuracy when regularization is applied to control for overfitting (held-out test set AUROC = 0.94; resampled test set AUROC = 0.974 ± 0.024 ; see *SI Appendix*, Fig. S2 and Table S1, for test set performance across algorithms). This finding supports the theory that transport capability is sufficient to determine whether a molecule is odorous or odorless, although additional experiments would be needed to validate that success

or failure in the obligatory steps of transport is sufficient to separate olfactory stimuli from odorless molecules.

Our transport ML model achieves high accuracy with few features, but experimental values of BP, vapor pressure, and log P are unavailable for many molecules. We also developed a simple rule of thumb that sacrifices some accuracy but can be applied knowing only the molecular formula (e.g., $\text{C}_{10}\text{H}_{12}\text{O}_2$): the “rule of three” states that molecules with molecular weight between 30 and 300 Da and with fewer than three heteroatoms are generally odorous (Fig. 1D). In our dataset, 96% of the molecules that meet these criteria are odorous.

Next, we asked if the relative accuracy of the transport ML, many-feature ML, and rule of three models varied by chemical class. Fig. 1E shows the test set AUROC for common chemical classes, averaged over 80 XGB models trained on randomized train/test splits. All models achieved strong predictive accuracy for alkanes, alcohols, and carbonyl-containing molecules, but only ML models accurately classified organohalides. The rule of three underperforms on inorganic compounds (e.g., NaCl) because they more often have ionic bonds, which dramatically raises the BP for a given molecular weight. The strong performance of the transport XGB model independent of chemical class supports the reliability of the model across common classes of molecules but also shows that simple rules of thumb (e.g., rule of three) may have limits to their applicability.

Importance of High-Quality Experimental Data. Why was this straightforward principle not established until now? Building a high-performing transport model required correction of two major sources of error: BP values and odor classifications. BPs reported in compendia or by chemoinformatic software are commonly estimated from chemical formula (8, 9), but these estimates are error-prone (median absolute error of 55 °C via Burnop method and 37 °C via Banks method) (Fig. 2A). Using such estimated BP values in an otherwise identical ML transport model worsens the classification performance (Fig. 2B and C: CV AUROC = 0.93 [Fig. 2B] and 0.95 [Fig. 2C]; error rate = 6.0% [Fig. 2B] and 4.5% [Fig. 2C]) compared to our final ML transport model (Fig. 1B: CV AUROC = 0.99; error rate = 2.6%). We replaced estimated BP with experimentally derived values (EPI Suite, US EPA) in our dataset and used estimates only when no experimental values were available; our final dataset includes experimental BP values for 1,270 molecules (66%) and either experimental BP or vapor pressure values for 1,408 molecules (73%) (Dataset S1). Experimental values should thus be used to generate high-confidence odorous/odorless predictions, although estimated values may be adequate where a higher error rate is acceptable.

The second and more important major source of error that we rectified is the odor classifications themselves. ML models can tolerate some noise in the training data, but inaccuracies in the test set can be more costly; specifically, predictive performance is bounded from above by mislabeled data in the test set (10). To ensure our test set was composed of accurately classified molecules, we supplemented our curated literature and web-scraped data ($n = 1,799$) with 128 additional molecular odor classifications (111 initially labeled odorous and 17 initially labeled odorless) generated through human psychophysics experiments. We analyzed all 111 stimuli that had a human-detectable odor using paired gas chromatography–mass spectrometry (GC–MS) and GC–olfactometry (GC–O) as a quality control (QC) measure to identify cases in which impurities, rather than the target compound, were responsible for the odor detected by human subjects. We found that 22% of molecules

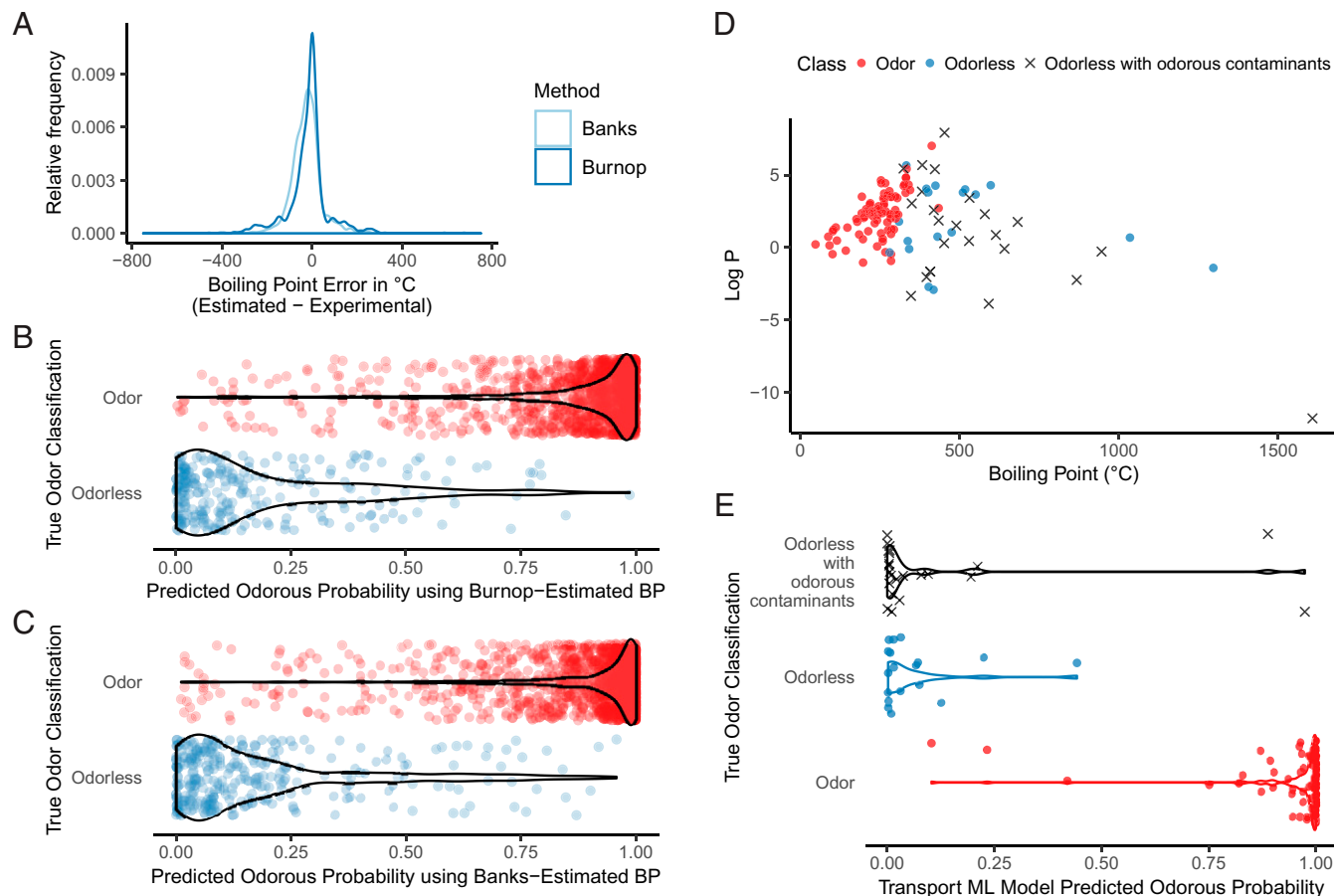


Fig. 2. Common inaccuracies in data impact model performance. (A) Difference between experimentally determined BP values and BP values calculated using the Burnop (9) and Banks (8) methods. (B and C) Odor classification predictions by transport-feature ML models using BP values calculated by the (B) Burnop or (C) Banks method. (D) Human subject-classified molecules in transport space defined by BP and log P. Many clearly nonvolatile molecules were initially classified as odors due to odorous contaminants. (E) Transport-feature ML model odor predictions for human subject-classified molecules. Chemical compounds that are odorless but had odorous contaminants are correctly predicted to be odorless by the model.

classified as odorous were actually odorless compounds contaminated with odorous compounds, despite high nominal purity ratings from vendors (Fig. 2D). Had we not performed this QC, we would have falsely believed that model performance was poor (precorrected transport ML AUROC = 0.81). In fact, most disagreements between our model's predictions and pre-QC classifications were due to the model correctly identifying mislabeled data (Fig. 2E).

Chemical compounds are common stimuli in olfaction research, but the impact of impurities on data is rarely discussed. Odor detection thresholds vary by many orders of magnitude across molecules, so even high purity (e.g., 99%) is insufficient grounds to consider odor to be driven entirely by the dominant molecular species (11). Addressing the impact of contaminants is thus vital for accurately measuring model performance.

Enumerating Odor Space. The dispute over the size of odor space has endured in the field because the criteria for odorous molecules were not rigorously defined. We now have a tool to address this debate: a simple quantitative model based on the well-developed field of physical transport (12) that makes highly accurate odor classification predictions and generalizes to molecules outside the training set. Chemical space is vast, but it can be enumerated. Applying graph theory and rational chemical constraints (e.g., bonds per atom, bond angles, and ring strain), Ruddigkeit et al. (13) developed GDB-17, a database of 166 billion unique molecules with heavy atom count (HAC) of

17 or fewer. While it excludes some known odorants (e.g., silica-containing molecules), the composition of this database of small organic molecules (limited to atoms C, H, N, O, S, F, Cl, Br, and I; HAC ≤ 17; unstable structures eliminated) makes it well matched to the goal of finding new odorous molecules.

We randomly down-sampled GDB-17 to create a representative subset of 107,086 molecules, including all unique structures for HAC 1 to 7 and ~10,000 molecules each for HAC 8 to 17. To this subsample, we applied our transport model to predict whether each molecule would be odorous. Fig. 3A shows that the proportion of molecules predicted to be odorous is highly dependent on HAC, so we applied the proportion of predicted odorous molecules per HAC in our subsample to estimate the number of odorous molecules with the same HAC in the full GDB-17 (Fig. 3B). Using this approach, of the 166 billion molecules in GDB-17, our transport model predicts 40 billion molecules will be odorous. Forty billion should be interpreted as a lower bound for the size of odor space because we know the odorous range extends beyond HAC 17 and that GDB-17 does not include all possibly odorous chemical structures.

The laws of physical transport upon which our model is built apply to all molecules. While most datasets used to study this question come from chemically biased flavor and fragrance databases, we intentionally purchased and classified the most chemically diverse set of molecules possible within the bounds of safety and availability. Because transport features were sufficient to classify molecules in our chemically diverse test set, we

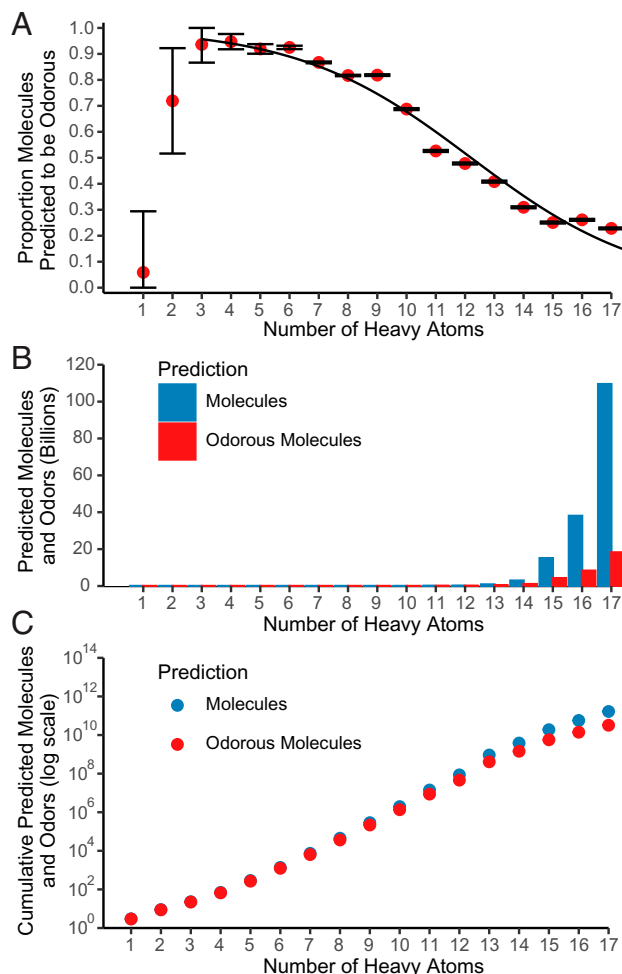


Fig. 3. The transport model can be used to predict the population of odor space. (A) Proportion of molecules predicted by the transport ML model to be odorous as a function of HAC. Red circles show the mean probability generated for HAC tranches from the GDB database (13) with SE indicated. (B) Estimated number of possible molecules and predicted odorous molecules from the GDB databases as a function of HAC. (C) Cumulative estimates of possible molecules and odorous molecules with increasing HAC on a logarithmic scale. The red data point at HAC 17 reflects our conservative estimate of 40 billion odorous molecules.

expect the model will successfully extrapolate to all of GDB-17. Even if we restrict our search to molecules in GDB-17 with neighbors in our training data (Tanimoto similarity of bit-based Morgan fingerprints > 0.4), we still find 10^7 unique predicted odorants (*SI Appendix, Fig. S3*), a value three orders of magnitude larger than the commonly cited estimate of 10,000 odorants (1).

Mapping Odor Space. Any database or catalog of purchasable odorous molecules is dwarfed by the scale of odor space. For example, the Sigma-Aldrich Flavor and Fragrance Catalog currently lists only 1,275 molecules. Many densely populated regions of odorous chemical space (Fig. 4A; interactive version available at <http://odormap.pyrfume.org/>) are sparsely represented by known odorants (Fig. 4B). Our map of the odorous region of chemical space can identify likely odorants (Fig. 4C and E) and be used to filter out unlikely odorants (Fig. 4D). The sheer number of as-yet-unsynthesized odorous molecules is striking, and there are whole classes of odorous molecules that have not been synthesized and whose odor characteristics are to date unknown.

Our results support the theory that drawing the borders of olfactory space is largely a matter of understanding physical transport. The accuracy of the transport model suggests that

nearly all molecules that can physically enter a receptor binding pocket have an odor. Although there are some clear exceptions, for some individuals, such as androstene (14), few molecules appear to be odorless simply because there is no sufficiently sensitive receptor. There are factors beyond volatility and hydrophobicity, such as interactions with transport proteins or enzymes in the mucosa (15, 16), that may influence whether and how a molecule smells. The success of our model shows that there is not much accuracy to gain by explicitly accounting for these factors, either because their impact on odor classification is relatively small or because the transport features we used are strongly correlated with some of these factors. Another piece of evidence is that the odor status of a molecule can change when environmental conditions change its transport capability: methane is odorless at standard pressure but reportedly smells of camphor when given to divers at 13 atmospheres (17). The receptor repertoire over the course of human evolution could not have evolved to detect this molecule, yet this repertoire maintains enough general sensitivity to—once given access to it—detect it nonetheless.

We aimed to collect a chemically diverse dataset but were limited by the types of molecules for which odor classifications are published, availability of molecules for purchase, and safety of molecules for human testing. There may be classes of odorous molecules that were not included in our training data and therefore not represented by our model. Our model was trained on human-made odor classifications, and so the resultant model should be understood to draw boundaries of human olfactory space. While we think it likely that transport features delineate olfactory space across species, differences in mucosal properties and receptor populations may shift these boundaries. Additional evidence that variables underlying olfactory perception are universal comes from Hahn et al. (3), who developed a theoretical model of odor intensity that agreed with experimental data from multiple species (humans, bullfrogs, and tortoises)

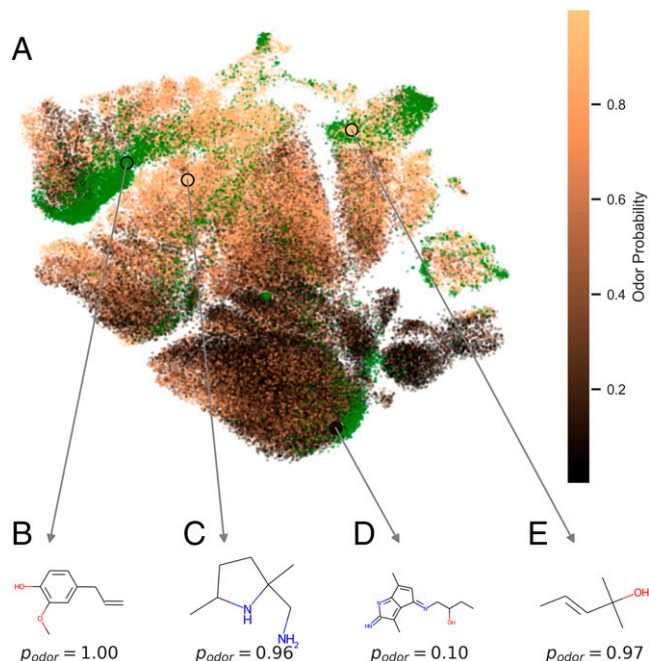


Fig. 4. Visualization of olfactory space highlights understudied regions. (A) UMAP plot of known odorous molecules (green) and possible molecules from GDB-17 colored by their transport ML-predicted odorous probability. Many regions dense with probable odors are sparsely represented by known odors. (B–E) Example molecules from GDB-17 and their transport ML-predicted probability of being odorous (p_{odor}).

after accounting for anatomical differences (i.e., length and thickness of the olfactory mucosa). Our model proposes an answer to the question of what makes a molecule odorous, but many questions remain. Our estimate of the number of possible odorous molecules cannot resolve the number of discriminable odor percepts (18–20); many of these predicted odorants may have indistinguishable odor percepts, and odorant mixtures may produce percepts that are distinct from that of any single odorant. Everything we know about odorants is derived from a tiny subset of all volatiles—a typical catalog of all volatiles present in foods likely represents less than 0.000003% of the molecules we can smell. This model invites researchers into the unknown, providing a map to uncharacterized regions of odor space and the means to representatively sample it.

Materials and Methods

Complete descriptions of methods are included in *SI Appendix, Materials and Methods*.

Building a Dataset for Model Training and Testing. Gathering a large set of clean data was a point of emphasis in this study. Our dataset includes 1,924 unique molecules labeled odorous or odorless (Dataset S1). We generated high-confidence classifications for 128 structurally diverse molecules through a three-alternative forced choice task with human subjects (75 trials per molecule; Dataset S2). The protocol was approved by the University of Pennsylvania Institutional Review Board, and 90 normosmic participants who consented to our study were tested. Following human “odorous” classification, we confirmed that the perceived odor was due to the nominal compound rather than a contaminant using paired GC–MS and GC–O. The remaining 1,796 were classified through literature and database searches. Molecules with ambiguous odor information (e.g., “practically odorless”) or discrepant classifications across different sources were thrown out. We calculated physicochemical descriptors for each molecule using Dragon (Taleté v6) and EPI Suite (US EPA), choosing experimental over estimated BP and vapor pressure values whenever possible.

Training Odor Classification Models. We initially set aside 30 molecules of the 128 with high-confidence, laboratory-derived classifications to form a test

set; the remaining 1,894 molecules were used to train models. We preprocessed data and trained models using the caret package (version 6.0.81) in R (version 3.5.3); the algorithm that optimized CV AUROC (extreme gradient boosting) was selected. We trained both complex (>3,000 physicochemical features; many-feature) and simple (3 features; transport) models and measured their classification accuracy on the held-out test set. We then ran an additional 25 splits containing different random sets of 30 (out of the 128) to generalize our results.

Estimating Number of Possible Odorants. We down-sampled the GDB-17 database to build a representative set of ~107,000 possible molecules with 17 or fewer heavy atoms (Dataset S3). To this dataset, we applied our transport model and generated a probability that each molecule would be odorous. Because odorous probability is heavily dependent on HAC, we stratified the molecules by HAC and calculated the proportion of molecules predicted to be odorous per HAC. We applied these proportions to the full GDB-17 to calculate a conservative estimate of the size of odor space. We visualized known odorants ($n = 8,366$; <https://pyrfume.org>) within the context of this newly defined odor space using uniform manifold approximation and projection (UMAP), available at <http://odormap.pyrfume.org>.

Data Availability. All code and data needed to train the models and reproduce the analyses shown in Figs. 1 and 2, as well as odor classifications and physicochemical features for 1,924 molecules, are publicly available via GitHub (<https://github.com/emayhew/OlfactorySpace>). There are no restrictions on the availability of these data. Physicochemical features for the 107,000 molecules sampled from GDB-17 are not included, due to terms of service from feature calculation software Dragon (Taleté). Molecular identifiers (simplified molecular-input line-entry system [SMILES]) for the 107,000 molecules are provided so that others can calculate features with a Dragon license. Datasets S1–S3 are also available on GitHub (<https://github.com/emayhew/OlfactorySpace/tree/main/tables>).

ACKNOWLEDGMENTS. We acknowledge support from the George Preti Research Support Core for Analytical Chemistry and thank George Preti, Katherine A. Prokop-Prigge, Bruce A. Kimball, Kai Zhao, and Gary K. Beauchamp for comments and discussions. This research was supported in part by grants from the NIH (Grants R01DC013339, U19NS112953, R01DC018455, and R01DC017757) and the Ajinomoto Co. Innovation Alliance Program. E.J.M. was supported by grants from the NIH (Grants T32DC000014 and F32DC019030).

1. L. Buck, R. Axel, A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
2. J.-N. Jaubert, C. Tapiero, J.-C. Dore, The field of odors: Toward a universal language for odor relationships. *Perfum. Flavorist* **20**, 1–16 (1995).
3. I. Hahn, P. W. Scherer, M. M. Mozell, A mass transport model of olfaction. *J. Theor. Biol.* **167**, 115–128 (1994).
4. H. Boelens, Structure-activity relationships in chemoreception by human olfaction. *Trends Pharmacol. Sci.* **4**, 421–426 (1983).
5. L. Ruddigkeit, M. Awale, J. L. Reymond, Expanding the fragrance chemical space for virtual screening. *J. Cheminform.* **6**, 27 (2014).
6. A. Sharma *et al.*, DeepOlf: Deep neural network based architecture for predicting odorants and their interacting olfactory receptors. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **19**, 418–428 (2020).
7. R. Gupta *et al.*, OdoriFy: A conglomerate of artificial intelligence-driven prediction engines for olfactory decoding. *J. Biol. Chem.* **297**, 100956 (2021).
8. W. H. Banks, Considerations of a vapour pressure-temperature equation, and their relation to Burnop's boiling-point function. *J. Chem. Soc.* 292–295 (1939).
9. V. C. E. Burnop, Boiling point and chemical constitution. Part 1. An additive function of molecular weight and boiling point. *J. Chem. Soc.* 826–829 (1938).
10. V. Gulshan *et al.*, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).
11. M. Paoli *et al.*, Minute impurities contribute significantly to olfactory receptor ligand studies: Tales from testing the vibration theory. *eNeuro* **4**, ENEURO.0070-17.2017 (2017).
12. E. McLaughlin, Transport properties of liquids in relation to their structure. *Q. Rev. Chem. Soc.* **14**, 236–256 (1960).
13. L. Ruddigkeit, R. van Deursen, L. C. Blum, J. L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
14. A. Keller, H. Zhuang, Q. Chi, L. B. Vosshall, L. Matsunami, Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–472 (2007).
15. M. Tegoni *et al.*, Mammalian odorant binding proteins. *Biochim. Biophys. Acta* **1482**, 229–240 (2000).
16. J. M. Heydel, P. Faure, F. Neiers, Nasal odorant metabolism: Enzymes, activity and function in olfaction. *Drug Metab. Rev.* **51**, 224–245 (2019).
17. P. Laffort, C. Gortan, Olfactory properties of some gases in hyperbaric atmosphere. *Chem. Senses* **12**, 139–142 (1987).
18. C. Bushdid, M. O. Magnusco, L. B. Vosshall, A. Keller, Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).
19. R. C. Gerkin, J. B. Castro, The number of olfactory stimuli that humans can discriminate is still unknown. *eLife* **4**, e08127 (2015).
20. M. Meister, On the dimensionality of odor space. *eLife* **4**, e07865 (2015).