ACS OMEGA

http://pubs.acs.org/journal/acsodf
Article

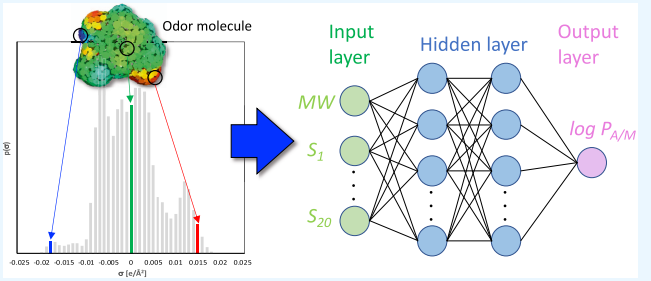
Prediction on Air−Nasal Mucus Partition Coefficients of Odor Compounds

Junwei Shen,* Yuki Harada, and Shinichiro Nakamura*

Cite This: *ACS Omega* 2024, 9, 31328−31334
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The air−nasal mucus partition coefficient is a crucial property among all of the interaction mechanisms between odor molecules and olfactory receptors, since this property contributes to our sense of smell. Due to the complexity of the mucus composition, in vivo determination of the air−mucus partition coefficient is a technical challenge. A predictable model of the air−mucus partition coefficient can provide valuable insights into the chemical properties that govern olfactory perception and can help design desired odorants. In this study, we propose a novel model based on the deep-layer neural network (DNN) algorithm to predict the air−mucus partition coefficients for a range of odor compounds. The molecular surface charge density (σ-profile) calculated from the COnductor like Screening MOdel for Real Solvents (COSMO-RS) thermodynamic package was adapted as descriptors of structural features of odor molecules. The results revealed that the air−mucus partition coefficients are highly correlated to the σ-profile of the studied compounds. The information obtained from the study provided interpretable results, which not only help in identifying the molecular features that contribute to the air−mucus partition coefficient of odorants but also aid in the design of compounds with the desired odor properties.



INTRODUCTION

Odor molecules inhaled into the nose must first dissolve in the mucus layer that lines the nasal cavity before they can reach the olfactory receptors.[1−4] The air−mucus partition coefficient of odorants is a crucial factor in determining the distribution of odorants within the olfactory region of the nose and ultimately influences the ability of the olfactory system to detect and distinguish different odorants. In general, a low air−nasal mucus partition coefficient means that the odorant tends to dissolve readily in the mucus layer and is likely to come into contact with the olfactory receptor cells that are embedded in the mucus. Consequently, there is a high likelihood of such an odorant being detected by the olfactory system.

Although in vivo determination of the air−mucus partition coefficient of odorants provides valuable insights into the olfactory system, the in vivo measurement of air−nasal mucus partition coefficients requires careful and specialized techniques. The chemical composition of nasal mucus is complex, typically containing a mixture of water, mucins, globular proteins, salts, carbohydrates, etc.[5−10] The properties of the mucus layer can vary depending on several factors, such as the location of the sample in the nasal cavity, the health status of the subject, and the adsorption of other substances in the mucus. It may be a source of significant error and uncertainty due to the complex and dynamic environment of the nasal mucus layer. Many researchers usually use the air−water ($P_{A/W}$) or octanol−water partition ($P_{O/W}$) coefficient to approximate the air−mucus partition coefficients. However, the air−mucus partition coefficient of odorants cannot be simply substituted by their log $P_{A/W}$ or log $P_{O/W}$ in most cases because of the complexity of the components in the mucus layer. For example, Mozell and Hornung[11] showed that the solubility of odorants in mucus or mucosa can be approximated by that in water for hydrophilic odorants; however, for hydrophobic odorants, the solubility in mucus or mucosa can be considerably higher than their aqueous solubility. They assumed that the difference may come from the interaction between odor molecules and the mucopolysaccharide and lipophilic carrier proteins contained in the mucus layer. Kurtz et al. investigated the odorant mucosal solubility and odorant diffusive transport in the mucosa by means of experimental and computational fluid dynamics (CFD) techniques.[12] They indicated that the transport of odor molecules in nasal mucosa clearly differs from that within an aqueous layer. These investigations suggest that the factors influencing odorant absorption/solubilization phenomena in the mucus layer are complex and multifactorial; there is not a

Received: October 5, 2023
Revised: March 22, 2024
Accepted: June 25, 2024
Published: July 9, 2024

straightforward way to deal with the air−mucus partition coefficients based on their log P values.

A predictable model of the air−mucus partition coefficient provides valuable insights into the chemical properties that govern olfactory perception. Therefore, it also helps to design the desired odorants. It gives an understanding of the environmental factors of odor recognition. A good predictable model usually relies on both the quantity and quality of the input data. In the case of air−mucus partition coefficients, there are multiple chemical properties that influence the partitioning behavior. There are only limited experimental data available due to the technical challenges. The lack of available data will make it difficult to develop accurate models for the air−mucus partition coefficients of odor molecules. By using reliable computational models presented previously, it is possible to obtain reliable estimates of the physical properties of odor molecules, particularly in cases where the experimental data are missing or incomplete.

The COnductor like Screening MOdel for Realistic Solvents (COSMO-RS) theory is an established method for predicting the thermodynamic properties of fluids.[13] In the COSMO-RS model, the molecular surface is divided properly in segments, and each segment has an average over one pixel charge density obtained from the electrostatic potential of the molecular electron density by means of quantum chemical calculations. A conductor screening charge density profile (so-called "σ-profile"), reducing the three-dimensional (3D) charge distribution to a two-dimensional (2D) histogram, characterizes the electrostatic polarity and charge distribution of the molecule of interest.

A representative example of the σ-profiles is shown in Figure 1 for nonpolar hexane, benzene, and polar water molecules.

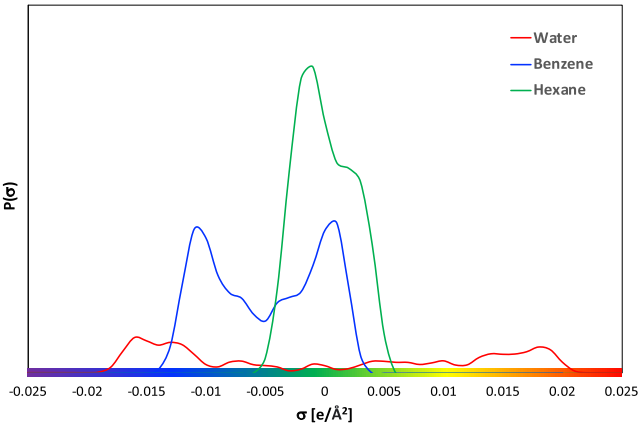


Figure 1. Three representative σ-profiles of nonpolar (hexane, benzene) and polar (water) molecules.

The entire σ-profile areas can be roughly divided into three regions: the electronic basicity region, the nonpolar region, and the electronic acidity region from the negative to positive-σ range. As illustrated in Figure 1, the σ-profile of the water molecule has a relatively broad distribution, with two almost symmetrical peaks in the negative and positive regions, attributed to two polar hydrogen and one oxygen atom, respectively. On the other hand, the nonpolar benzene molecule shows two symmetrical picks in the region slightly off the center, where the pick in the positive σ-region can be assigned to a delocalized π-electron in benzene. For hexane, there is a narrow distribution of nearly neutral charge density around zero, with a shoulder appearing in the slightly positive region arising from the carbon atoms of hexane. The σ-profiles, being capable of taking into account solvation behavior, are quite suitable for universal descriptors in the studies of quantitative structure−property relationship.[14−16] Notice that several studies of machine learning have been carried out by using molecular σ-profiles as descriptors in the prediction of various properties such as viscosity of ionic liquids[17] and heat capacity.[18]

To date, deep learning technology has been successfully applied to solve complex problems in many scientific areas of interest due to its ability to handle large amounts of data, capture nonlinear relationships, extract meaningful features, transfer knowledge across domains, and generalize to new data.[19] Over the last few decades, a number of deep learning models for predicting partition coefficients of organic compounds have been developed.[20,21] There are also several studies that have explored the use of deep learning methods for predicting the olfactory properties of molecules.[22−24] The deep neural network (DNN) is a mathematical model that functions as a universal approximator. The DNN and its associated algorithms are already successfully used in various chemoinformatic applications. In environmental risk assessment,[25−27] DNNs should have similar performance in the prediction of air−mucus partition coefficients of odor compounds.

In this work, we report an approach to a deep learning model for the prediction of the air−mucus partition coefficients of odorants. We emphasize the useful σ-profile as a descriptor. Figure 2 briefly depicts the overall framework of this work. First, the σ-profiles of odorants were directly collected from the precalculated COSMO-RS database or generated by using the quantum chemical density-functional-theory (DFT) calculations. The data sets of σ-profiles of odorants with molecular weights were divided into training and test sets. The DNN model was then trained and optimized by using the training data set. Since only little data is available in training sets, overfitting may become an issue, which leads to poor performance on test or prediction results. To prevent overfitting and improve the generalization performance of the model, the early stopping technique was adopted to monitor the validation loss during training and to stop it when the validation loss starts to increase before the overfitting of the training data. To ensure reliability and accuracy, the previously established model was validated with the same test data set. In addition, the accuracy of predicted results was compared with the traditional method using macroscopic property-based descriptors such as vapor pressure, Henry's law constant, and octanol−water partition coefficient.

COMPUTATIONAL DETAILS

The air−mucus partition coefficients of 66 odorants were taken from the estimated values of Scott et al.,[28] since there is only a small amount of experimental data available. They showed the correlation of a linear relationship of $\log P_{O/W}$ on the difference between $\log P_{A/M}$ and $\log P_{A/W}$ values with a regression of $R^2 = 0.81$ to the experimental data for 12 odor molecules,[12] where the $\log P_{O/W}$ and $\log P_{A/W}$ values used in the estimations were obtained through the U.S. Environment Protection Agency database.[29] In this work, we used these estimated values as target variables in training and validation.

The σ-profiles of 66 odor molecules were directly collected from the precalculated COSMO-RS database or generated by using DFT calculations at the same level (BP86/def2-

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334

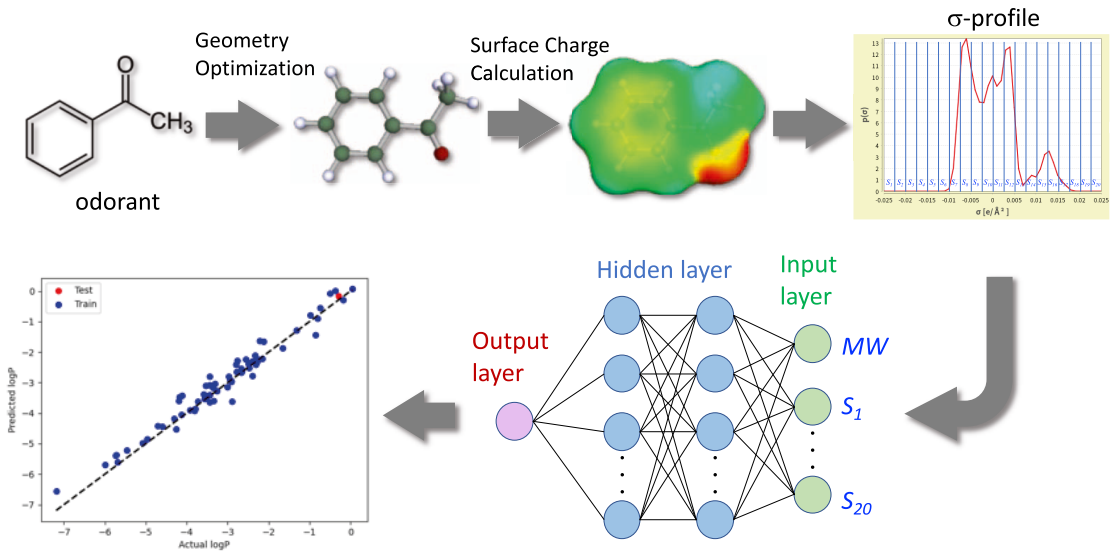


Figure 2. Deep learning scheme for the prediction of the air−mucus partition coefficient based on the σ-profile descriptors of odor molecules.

TZVP).[30] In several previous studies that used the σ-profiles as descriptors, the σ-profile range (−0.025 to +0.025 e/Å^2) was usually divided into 6 or 10 equal segments (e.g., Zhou et al.).[15] However, we anticipate that a low resolution of the segmentation might result in the loss of electrostatic characteristics of the molecule. After examining the dependence of the σ-profile segmentation size on model performance, we divided the entire σ-profile range of the molecule into 20 areas to present molecular features accurately. Thus, a total of 20 feature vectors (S1, S2, ···, S20) were obtained for each odor molecule by integrating those regions. Afterward, a min−max normalization process was applied to these σ-profile descriptors to ensure that all feature variables have the same scale, and then we added one more element of the molecular weight to each molecule.

The DNNs can provide an efficient architecture to predict a variety of physical properties. The optimal number of hidden layers and neurons in a DNN model is usually dependent on the property of the problem and the size and complexity of the data set. In most cases, one or two hidden layers are sufficient to handle the prediction of the compound properties. In this study, a DNN architecture (Keras model, python3.9) was designed with two hidden layers. The number of neurons was set as 100 in each hidden layer, and the rectified linear unit function was adopted for the activation function of the hidden layers. The stochastic gradient-based optimizer (adaptive moment estimation, "adam") was used as a solver for weight optimization, and the maximum number of solver iterations was set to 1000 epochs. The cross-validation techniques monitoring the mean square error (MSE) were applied to evaluate the generalization performance of the model.

To examine and evaluate the performance of the thus-obtained DNN model, the predicted results were compared with the traditional approach that uses macroscopic property-based descriptors (such as Henry's law constant, octanol−water partition coefficient, etc.) instead of σ-profiles. These thermodynamic properties of 66 odorants were also predicted by the COSMO-RS approach and validated with the available experimental data.

RESULTS AND DISCUSSION

To obtain an unbiased estimation of the model performance, we need to evaluate it with the data independent of training sets. In this work, two cross-validation techniques were applied for model and parameter selection and accuracy estimation. The first one is the hold-out method, which is the simplest kind of cross-validation to evaluate a regression. It randomly partitions the 66 data sets into two subsets (53 training and 13 validation data sets) in a given ratio of 8:2. Another is the leave-one-out (loo) cross-validation, in which at a time only one test set input datum is used, while the rest 65 data are used to train the model; thus, the validation was performed over all data sets.

Figure 3 shows the training and testing results of a DNN model using a hold-out (split ratio 8:2) validation method for

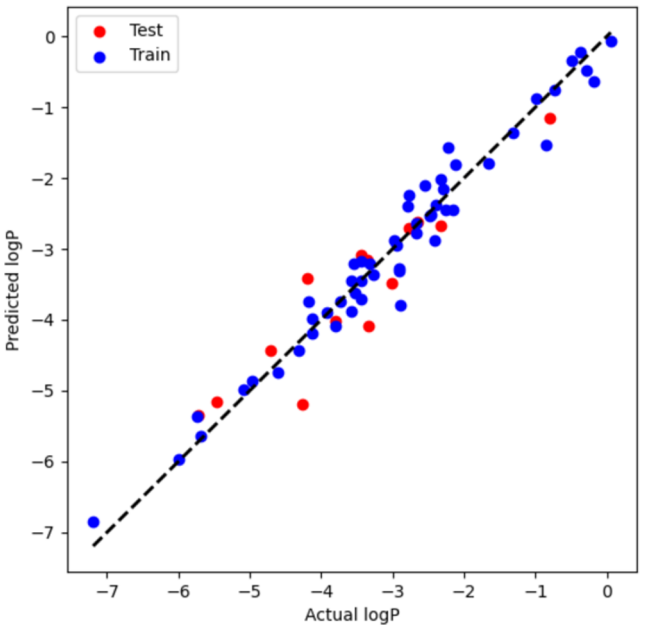


Figure 3. Example of the training and test results of a DNN model using hold-out validation technique (R^2 = 0.936 ± 0.017).

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334

the air−mucus partition coefficients based on the σ-profiles of odorants. The correlation coefficient for the model with nested/repeated 10-fold cross-validation (10 different splits of the data into 10 independent cross-validation testing sets) is 0.936 ± 0.017. A representative curve for the loss function as a function of the epoch number in training and testing the model is shown in Figure 4. It is clearly shown that both training and

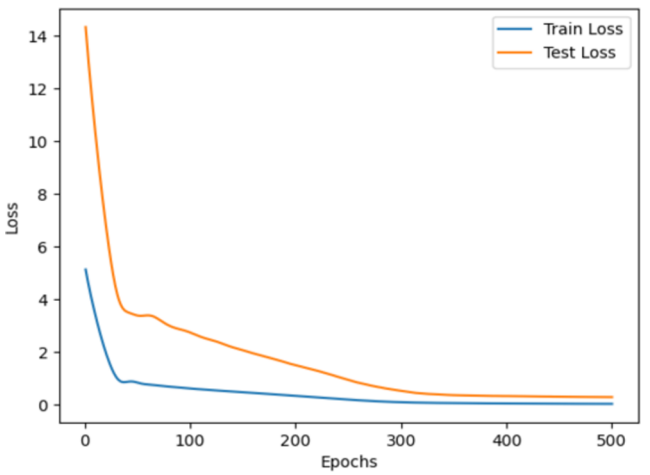


Figure 4. Example of the loss function for the model with training and test data iterating over each epoch.

test curves of loss function declined rapidly and flattened out at approximately 300 epochs. It indicates that the model has a fast convergence and high stability.

The results of loo validation are shown in Figure 5. The spots are shown in red (test) and in blue (train) at each time. Notice that the blue spots are the best correlation result, whereas the red spots are the values for each validation. Taking an average over each test run, we obtained a correlation of 0.957 ± 0.019.

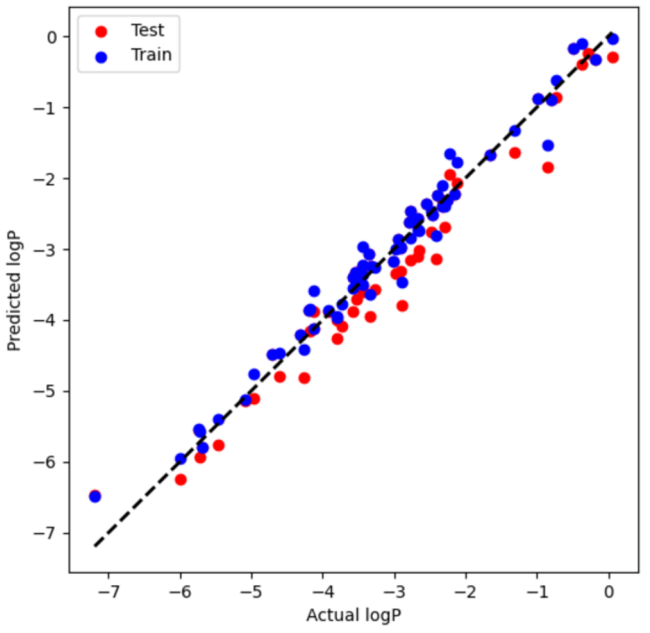


Figure 5. Best training result (blue dots) with test result (red dots) of a DNN model using leave-one-out validation technique (R^2 = 0.957 ± 0.019).

According to the results for two validations, it is shown that the loo method seems to be more appropriate than hold-out for the current study, since the loo can use more training data in each iteration. It enables the model to become better representations of the air−mucus partition coefficients. The best result predicted by the loo method is shown in Figure 6.

We performed the quantitative analysis of the prediction performance by using the training values to understand those that are important and critical features of odor molecule on its air−mucus partition coefficients. Permutation feature impor-

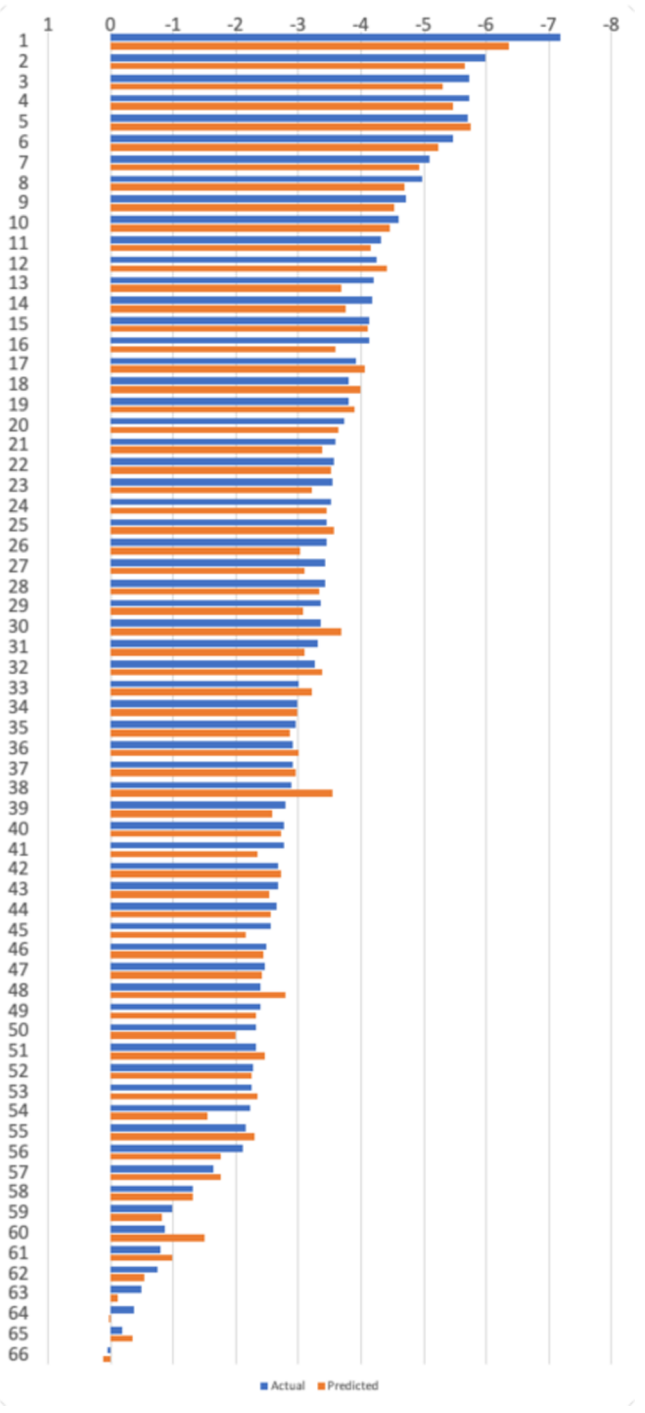


Figure 6. Comparison of predicted air−mucus partition coefficients with the corresponding literature data for 66 odorants used in this study.

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334

tance (PFI) algorithm based on Fisher et al.[31] has been frequently used in deep learning study, which measures the increase in prediction error caused by reordering feature values, thereby breaking the relationship between the feature and the outcome. A PFI plot of the current model is shown in Figure 7. As shown in Figure S1 in the SI, it is clear that

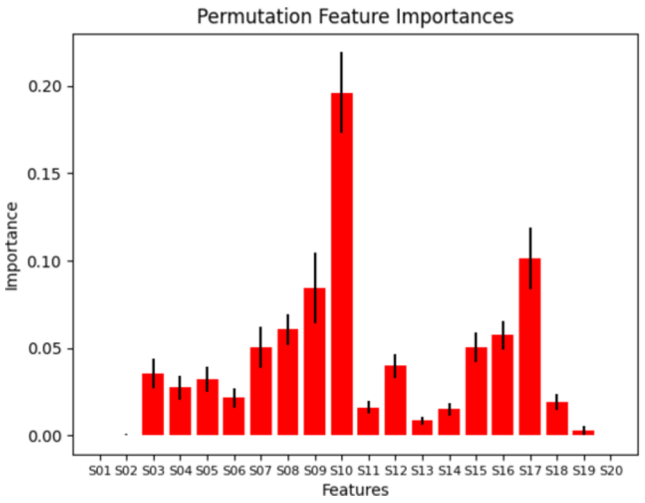


Figure 7. Permutation feature importance analysis for σ-profile descriptors in the DNN model with leave-one-out cross-validation.

molecular weight has a dominant influence on the predicted results. We leave out the molecular weight (M_W) from Figure 7, since it is due to the fact that the large molecules are less volatile and more difficult to dissolve into nasal mucus than small molecules.

Hereafter, we discuss the remaining elements between S1 and S20. The distribution in neutral (S10, ~0 e/Å^2) and positive screen charge (S17, ~0.017 e/Å^2) areas of odorants have a relatively larger contribution among air−mucus partition coefficients. The neutral one (S10) must suggest that the characteristic strong interaction between odor molecules and mucus components mainly consists of van der Waals relative to electrostatic interaction. Then, S17 suggests that somehow polarization must be included, to be discussed below.

To further explore the relationship between the molecular structure and the air−mucus partition coefficients, the surface charge characteristics of several representative odor molecules were analyzed in detail. As shown in Figure 8, the σ-profile of vanillin (black line) has a relatively symmetric distribution in both hydrophobic and hydrophilic regions due to the cumulative effect of its functional groups of aldehyde, hydroxyl, and ether. The vanillin with a moderate polarization degree thus resulted in the highest solubility in mucus ($\log P_{A/M}$ = −7.191) among the studied compounds. By comparing the peak position in the σ-profile, although hexanoic acid (blue line) has a more positive distribution and *tert*-butanol (red line) has more negative distribution than that of vanillin, the air−mucus partition coefficients, $\log P_{A/M}$, were −5.087 for hexanoic acid and −3.436 for *tert*-butanol. Both suggest much smaller solubility than that of vanillin (Table S1 in SI). The nonpolar molecule *n*-hexane, having only one distribution peak (green line) in the hydrophobic region, exhibits poor solubility in mucus (a higher $\log P_{A/M}$ = 0.053). The above-analyzed results reveal that the odorant with a balanced and moderate

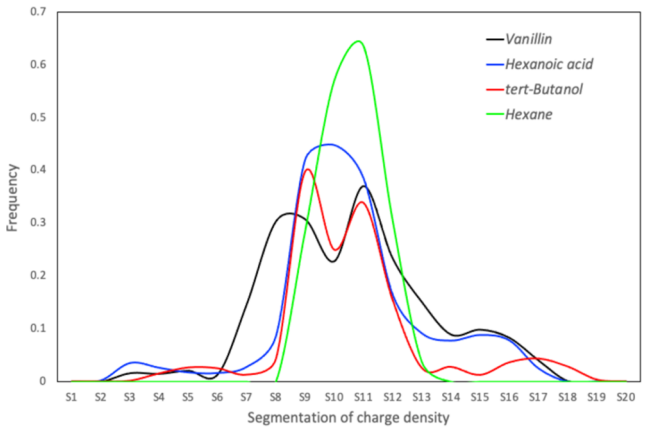


Figure 8. σ-Profiles of four representative odor molecules: vanillin, hexanoic acid, *tert*-butanol, and hexane.

positive/negative polarization tends to make a strong interaction with mucus components, leading to being more soluble in the mucus (i.e., a lower air−mucus partition coefficient) than others.

In order to evaluate the performance of the current DNN model, the predicted results were compared to the traditional approach that uses macroscopic property-based descriptors such as Henry's law constant and octanol−water partition coefficient. Because of the missing and limited data, the predicted values of Henry's law constants and octanol−water partition coefficients for these odorants were used in multiple linear regression analysis. Many studies have demonstrated that quantum-chemically based COSMO-RS theory can provide accurate predictions for partition properties.[13,32,33] The correlation of COSMO-RS predicted values of Henry's law constants (air−water partition coefficients) and octanol−water partition coefficients versus available experimental data for the studied odorants is shown in Figure S2 in SI. The coefficient of the determination for $\log K_H$ is R^2 = 0.9328 and that for $\log P_{O/W}$ is R^2 = 0.9324. The results of the linear regression of the air−mucus partition coefficients with Henry's law constants and octanol−water partition coefficients are shown in Figure 9. We obtained a linear regression with a correlation coefficient of R^2 = 0.7653.

$$\log P_{A/M} = -0.0138M_W + 0.1628\log K_H + 0.5474\log P_{O/W} + 0.3239$$

As compared to the correlation coefficients by the DNN model discussed above, it is obvious that the neural network model based on the σ-profile performed better than the traditional multiple linear regression equation on the prediction of the air−mucus partition coefficients of odor compounds. A comprehensive comparison of the estimated results using the Scott et al. equation, our DNN model, and the model based on the COSMO-RS prediction of the macroscopic properties for 12 odor molecules with the experimental values of the air−mucus partition coefficient is shown in Table S2 in SI.

The high solubility in mucus should be a fundamentally important and sensitive property for the smell. Notice that the model presented here evaluates the degree of dissolution of odorants in mucus but still not the odor intensity itself. When we use an experimental index (threshold)[34] for odor intensity evaluation as the target variables in the model training, it is

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334

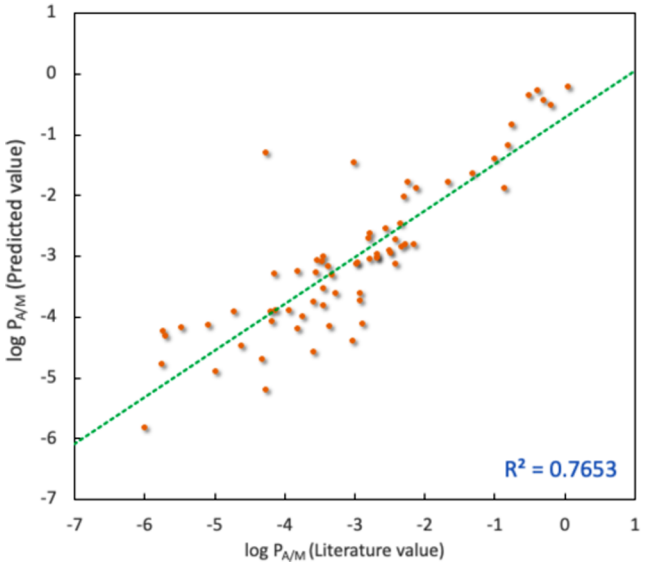


Figure 9. Accuracy of a multiple linear regression model for predicting the air−mucus partition coefficients using macroscopic property-based descriptors.

indispensable to consider the solubility before we get into the correlation between the odor intensity and structural characteristics of odorants. It is evident that the higher the partition of an odorant in the mucus layer, the higher the probability of contact between the odor molecule and the olfactory receptor.

CONCLUSIONS

In this article, a DNN model learns to represent the air−mucus partition behavior based on the surface charge density (σ-profile) descriptors of the odor molecules. The σ-profiles of molecules were computed from quantum-chemistry-based COSMO-RS solvation models. The results showed that the air−mucus partition coefficients are highly correlated with the σ-profiles of the odorants studied. Compared with traditional approaches that use macroscopic property-based descriptors, there was a significant improvement in the prediction accuracy of the air−mucus partition coefficients of odor compounds. The analysis of the influencing factors indicates that molecular weight is the dominant factor over a set of explanatory variables. This is natural because it is believed to be linked to the volatility of odor molecules. On the other hand, the molecular surface charge distribution is directly related to the intermolecular interaction between the odorants and the mucus. An analysis of the odor air−mucus partition behavior with the corresponding distribution of σ-profile reveals that the odorants with a balanced and moderate polarization favorably induce the interaction with mucus components, leading to being soluble in mucus. The information obtained from the study provided interpretable results, allowing us to generate hypotheses about how the molecular features contribute to the air−mucus partition coefficient of odorants. This can also help in understanding the chemical properties that govern olfactory perception, which will aid in the design of molecules with the desired odor properties. Furthermore, this study demonstrates the potential to build advanced machine learning applications for predicting various thermodynamic properties of volatile organic compounds, such as absorption rates of nasal drug agents and lethal doses of volatile toxicants, based on the molecular σ-profile information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c07722.

Permutation feature importance analysis for all descriptors used in the DNN model with the leave-one-out cross-validation method; the correlation of COSMO-RS prediction of (a) Henry's law constants and (b) octanol−water partition coefficients with the corresponding experimental data; comparison of the calculated air−mucus partition coefficients with the corresponding literature data for 66 odorants used in this study; comparison of various calculated air/mucus partition coefficients with the corresponding experimental data for 12 odorants (PDF)

AUTHOR INFORMATION

Corresponding Authors

Junwei Shen − *Laboratory for Data Sciences, Priority Organization for Innovation and Excellence, Kumamoto University, Kumamoto 860-8555, Japan;* orcid.org/0000-0003-4223-6735; Email: jwshen@kumamoto-u.ac.jp

Shinichiro Nakamura − *Laboratory for Data Sciences, Priority Organization for Innovation and Excellence, Kumamoto University, Kumamoto 860-8555, Japan;* Email: shindon@kumamoto-u.ac.jp

Author

Yuki Harada − *Laboratory for Data Sciences, Priority Organization for Innovation and Excellence, Kumamoto University, Kumamoto 860-8555, Japan;* orcid.org/0009-0000-4254-7803

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07722

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The present work is supported by The Chemo-Sero-Therapeutic Research Institute (KAKETSUKEN), Japan. The authors are grateful to professor emeritus Kunio Tanabe of the Institute of Statistical Mathematics for his valuable advice and meaningful discussion on deep learning methods. The authors also gratefully acknowledge Prof. Hirofumi Kai, of Kumamoto University, for his valuable advice and fruitful discussions.

REFERENCES

(1) *The Neurobiology of Olfaction*; Menini, A., Ed.; CRC Press/Taylor & Francis: Boca Raton (FL), 2010.

(2) *The Human Sense of Smell*; David, G. L.; Richard, L. D.; Winrich, B., Eds.; Springer-Verlag, 1991.

(3) Shirai, T.; Takase, D.; Yokoyama, J.; Nakanishi, K.; Uehara, C.; Saito, N.; Kato-Namba, A.; Yoshikawa, K. Functions of human olfactory mucus and age-dependent changes. *Sci. Rep.* **2023**, *13*, No. 971.

(4) Amoore, J. E. Effects of Chemical Exposure on Olfaction in Humans. In *Toxicology of the Nasal Passages*; Barrow, C. S., Ed.; McGraw-Hill International Book Company: WA, 1984; pp 155−190.

(5) Allen, A.; Flemstrom, G.; Garner, A.; Kivilaakso, E. Gastroduodenal mucosal protection. *Physiol. Rev.* **1993**, *73*, 823−857.

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334

(6) Bansil, R.; Turner, B. S. Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interface Sci.* **2006**, *11*, 164−170.
(7) Button, B.; Okada, S. F.; Frederick, C. B.; Thelin, W. R.; Boucher, R. C. Mechanosensitive ATP Release Maintains Proper Mucus Hydration of Airways. *Sci. Signaling* **2013**, *6*, No. ra46, DOI: 10.1126/scisignal.2003755.
(8) Carlstedt, I.; Sheehan, J. K. Structure and macromolecular properties of cervical mucus glycoproteins. *Symp. Soc. Exp. Biol.* **1989**, *43*, 289−316.
(9) Cone, R. A. *Mucus, Mucosal Immunology*; Lamm, M. E.; McGhee, J. R.; Bienenstock, J.; Mayer, L.; Strober, W., Eds.; Academic Press: Burlington, 2005; pp 49−72.
(10) Thornton, D. J.; Sheehan, J. K. From mucins to mucus: toward a more coherent understanding of this essential barrier. *Proc. Am. Thorac. Soc.* **2004**, *1*, 54−61.
(11) Hornung, D. E.; Youngentob, S. L.; Mozell, M. M. Olfactory mucosa/air partitioning of odorants. *Brain Res.* **1987**, *413*, 147−154.
(12) Kurtz, D. B.; Zhao, K.; Hornung, D. E.; Scherer, P. Experimental and Numerical Determination of Odorant Solubility in Nasal and Olfactory Mucosa. *Chem. Senses* **2004**, *29*, 763−773, DOI: 10.1093/chemse/bjh079.
(13) Klamt, A. *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, 2005. Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem. A* **1995**, *99*, 2224−2235. Klamt, A.; Eckert, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43−72.
(14) Abranches, D. O.; Zhang, Y.; Maginn, E. J.; Colón, Y. J. Sigma profiles in deep learning: towards a universal molecular descriptor. *Chem. Commun.* **2022**, *58*, 5630−5633, DOI: 10.1039/d2cc01549h.
(15) Zhou, T.; McBride, K.; Zhang, X.; Qi, Z.; Sundmacher, K. Integrated Solvent and Process Design Exemplified for a Diels−Alder Reaction. *AIChE J.* **2015**, *61*, 147−158, DOI: 10.1002/aic.14630.
(16) Abranches, D. O.; Zhang, Y.; Maginn, E. J.; Colón, Y. J. Sigma profiles in deep learning: towards a universal molecular descriptor. *Chem. Commun.* **2022**, *58*, 5630−5633, DOI: 10.1039/D2CC01549H.
(17) Kang, X.; Zhao, Z.; Qian, J.; Afzal, R. M. Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm. *Ind. Eng. Chem. Res.* **2017**, *56* (39), 11344−11351, DOI: 10.1021/acs.iecr.7b02722.
(18) Kang, X.; Liu, X.; Li, J.; Zhao, Y.; Zhang, H. Heat capacity prediction of ionic liquids based on quantum chemistry descriptors. *Ind. Eng. Chem. Res.* **2018**, *57* (49), 16989−16994, DOI: 10.1021/acs.iecr.8b03668.
(19) Ching, T.; Himmelstein, D. S.; Beaulieu-Jones, B. K.; Kalinin, A. A.; Do, B. T.; Way, G. P.; Ferrero, E.; Agapow, P.-M.; Zietz, M.; Hoffman, M. M.; Xie, W.; Rosen, G. L.; Lengerich, B. J.; Israeli, J.; Lanchantin, J.; Woloszynek, S.; Carpenter, A. E.; Shrikumar, A.; Xu, J.; Cofer, E. M.; Lavender, C. A.; Turaga, S. C.; Alexandari, A. M.; Lu, Z.; Harris, D. J.; DeCaprio, D.; Qi, Y.; Kundaje, A.; Peng, Y.; Wiley, L. K.; Segler, M. H. S.; Boca, S. M.; Swamidass, S. J.; Huang, A.; Gitter, A.; Greene, C. S. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **2018**, *15*, No. 20170387, DOI: 10.1098/rsif.2017.0387.
(20) Win, Z.-M.; Cheong, A. M. Y.; Hopkins, W. S. Using machine learning to predict partition coefficient (Log P) and distribution coefficient (Log D) with molecular descriptors and liquid chromatography retention Time. *J. Chem. Inf. Model.* **2023**, *63*, 1906−1913, DOI: 10.1021/acs.jcim.2c01373.
(21) Zhu, Q.; Jia, Q.; Liu, Z.; Ge, Y.; Gu, X.; Cui, Z.; Fan, M.; Ma, J. Molecular partition coefficient from machine learning with polarization and entropy embedded atom-centered symmetry functions. *Phys. Chem. Chem. Phys.* **2022**, *24*, 23082−23088.
(22) Chacko, R.; Jain, D.; Patwardhan, M.; Puri, A.; Karande, S.; Rai, B. Data based predictive models for odor perception. *Sci. Rep.* **2020**, *10*, No. 17136.
(23) Shang, L.; Liu, C.; Tomiura, Y.; Hayashi, K. Machine-Learning-Based Olfactometer: Prediction of Odor Perception from Physicochemical Features of Odorant Molecules. *Anal. Chem.* **2017**, *89*, 11999−12005.
(24) Nozaki, Y.; Nakamoto, T. Predictive modeling for odor character of a chemical using machine learning combined with natural language processing. *PLoS One* **2018**, *13*, No. e0208962.
(25) Saini, V.; Kumar, R. A machine learning approach for predicting the empirical polarity of organic solvents. *New J. Chem.* **2022**, *46*, 16981−16989, DOI: 10.1039/D2NJ02513B.
(26) Zang, Q.; Mansouri, K.; Williams, A. J.; Judson, R. S.; Allen, D. G.; Casey, W. M.; Kleinstreuer, N. C. In Silico Prediction of Physicochemical Properties of Environmental Chemicals Using Molecular Fingerprints and Machine Learning. *Chem. Inf. Model.* **2017**, *57*, 36−49, DOI: 10.1021/acs.jcim.6b00625.
(27) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154−6168, DOI: 10.1039/C9CS00786E.
(28) Scott, J. W.; Sherrill, L.; Jiang, J.; Zhao, K. Tuning to Odor Solubility and Sorption Pattern in Olfactory Epithelial Responses. *J. Neurosci.* **2014**, *34*, 2025−2036, DOI: 10.1523/JNEUROSCI.3736-13.2014.
(29) *U.S. Environment Protection Agency Database*, https://www.epa.gov/oppt/exposure/pubs/episuite.html. (accessed September 10, 2023).
(30) *BIOVIA, COSMObase, TURBOMOLE*; Version 2023, Dassault Systèmes: San Diego, 2023.
(31) Fisher, A.; Rudin, C.; Dominici, F. Model Class Reliance: Variable importance measures for any machine learning model class, from the 'Rashomon' perspective. *J. Mach. Learn. Res.* **2019**, *20* (177), 1−38.
(32) Glüge, J.; Bogdal, C.; Scheringer, M.; Buser, A. M.; Hungerbühler, K. Calculation of Physicochemical Properties for Short- and Medium-Chain Chlorinated Paraffins. *J. Phys. Chem. Ref. Data* **2013**, *42*, No. 023103.
(33) Endo, S.; Hammer, J. Predicting Partition Coefficients of Short-Chain Chlorinated Paraffin Congeners by COSMO-RS-Trained Fragment Contribution Models. *Environ. Sci. Technol.* **2020**, *54*, 15162−15169.
(34) ASTM - E554. *Recommended Practice for Odor Suprathreshold Intensity Referencing* American Society for Testing Materials: Philadelphia, P.A., 1975.

https://doi.org/10.1021/acsomega.3c07722
ACS Omega 2024, 9, 31328−31334