Feature attributions for water-solubility predictions obtained by artificial intelligence methods and chemists

Teruhisa Sadakane, Koki Nakata, Kayo Suda, and Daisuke Yokogawa*

Graduate Schoool of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

E-mail: c-d.yokogawa@g.ecc.u-tokyo.ac.jp

Abstract

Recently, the field of explainable artificial intelligence has attracted significant research interest, with a particular focus on "feature attribution" in the field of chemistry. However, studies comparing the relationship between artificial-intelligence- and human-based feature attributions when predicting the same outcome are scarce. Hence, the current study aims to investigate this relationship by comparing machine-learning-based feature attributions (graph neural networks and integrated gradients) with those of chemists (Hansch–Fujita method) when predicting water solubility. The findings reveal that the artificial-intelligence-based attributions are similar to those of chemists despite their distinct origins.

Introduction

In recent years, advancements in neural network technology have spurred significant research interest in explainable artificial intelligence.[1–5] In particular, studies focusing on "feature at-

tribution," which analyzes the contributions of certain descriptors to the prediction results, have begun gaining traction in the field of chemistry. For instance, Kim et al.[6] clarified the influence of molecular structure coordinates as descriptors on the reaction coordinate of transition state structures in differentiation reactions by employing methods such as adapting local interpretable model-agnostic explanations[7] and SHaplay Additive exPlanations.[8] Similarly, Okuno et al. successfully developed a retrosynthetic reaction prediction system using integrated gradients (IGs), enabling the visualization of atoms pertinent to the reaction.[9]

Chemists have also proposed alternative concepts similar to feature attribution. Among these, the "substituent" or "functional group" concept is particularly notable. This concept was first proposed in the 19th century based on "the radical theory"[10–12] and "the theory of type"[13–16] through experiments on various chemical reactions.[17] Later, in the early 20th century, Hammett et al. established a method to theoretically predict hydrolysis reactions using substituents,[18] thus elucidating the contributions of substituents to such reactions. Consequently, substituents represent an important concept that originated from the link between chemists' experimental observations and theoretical frameworks.

The fields of machine learning and chemistry share similar concepts known as "feature attribution". However, owing to the distinct development periods and backgrounds of these concepts, as well as their differing prediction targets, comparative studies focusing on these concepts are scarce, at least to the best of our knowledge. Hence, if these concepts were capable of predicting the same physical property, would the feature attributions of machine-learning methods and chemists exhibit similarity? To answer this question, herein, we compare the feature attributions for water-solubility prediction acquired through graph neural networks (GNNs) and IGs with those obtained through the Hansch–Fujita method. Evidently, water solubility is a highly invariant and a crucial physical used in a wide range of applications such as designing imaging molecules,[19–21] drugs[22–24] and related compounds. Thus, the discussion of feature attribution in this investigation will prove significant for such design applications.

Methods

Model and descriptors

In this study, we employed the GNNs architectures proposed in our previous study.[25] The model comprised four components: convolution and concatenation layers, fully connected layers in each atom, a pooling layer to construct molecular features, and fully connected layers to obtain target properties. Among these, for the convolution and concatenation layers, we employed two descriptor sets: molecular-atomic properties (MAPs) and isolated-atomic properties (IAPs).[26] As indicated in Table 1, the MAPs represent atomic properties extracted from molecular calculations, whereas the IAPs represent atomic properties defined in each isolated atom.

Table 1: Descriptors employed in this study

molecular-atomic properties (MAPs)	isolated-atomic properties (IAPs)
constrained spatial electron density distribution charge	effective charge
partial Fukui function (+)	atomic polarization
volume	atomic radius
atomic dispersion coefficient	ionization energy
fractional anisotropy of the magnetic shielding constant	electron affinity
absolute value of the effective atomic orbital energy	mass
accessible surface area	accessible surface area
atomic fluctuation	atomic fluctuation

Datasets

In this study, we selected water solubility as the basis for comparing feature attributions. This choice was motivated by its advantageous features in machine learning and chemistry. Moreover, the abundant experimental data on water solubility is expected to facilitate the construction of accurate learning models. Moreover, chemists have also acknowledged the relationship between solubility and substituents.

The original dataset for this study was obtained from the dataset provided by DeepChem.[27]

From this dataset, we filtered out molecules containing more than 35 non-hydrogen atoms or ions. Ultimately, the modified dataset contained 1027 molecules, and this dataset was then divided into training and test datasets at a ratio of 8:2.

Definition of feature attribution

In the field of chemistry, researchers have proposed numerous methods defined on substituent constants to explain the physicochemical properties of molecules. The Hansch–Fujita approach is one such method, designed to elucidate the water-octanol partition coefficient (log P) of a molecule by focusing on the constants defined on the substituents.[28] Within the model framework, the substituent constant π_X is determined using the log P value of matrix benzene (H) and the substituent (X), as follows:

$$\pi_{\mathrm{X}} = \log P_{\mathrm{X}} - \log P_{\mathrm{H}} \tag{1}$$

If π_{X} is negatively (positively) large, the substituent (X) makes the molecule more hydrophilic (hydrophobic). From this character, we can use π_{X} as the attribution to explain the molecular solubility. Here, when π_X is significantly negative (Positive), the substituent (X) imparts increased hydrophilicity (hydrophobicity) to the molecule. Leveraging this, we can use π_X as an attribution to explain molecular solubility.

In a previous study, we proposed another attribution utilizing the GNNs and IGs technique.[25] The proposed attribution $\bar{G}_{in}$, was defined for atomic feature i and atomic site n. In this study, the attribution for the substituent X is defined as follows:

$$\boldsymbol{G}_X = \sum_{in \in \mathrm{X}} \overline{G}_{in} \tag{2}$$

Results and discussions

The primary purpose of this study was to assign attributions for determining the atomic groups crucial for hydration. However, given that the assignment was performed for the trained model, validating the reliability of the learning model before engaging in the attribution assignment discussion was crucial. Table 2 summarizes the root mean squared error (RMSE) and coefficient of determination (R^2) values for the prediction of water solubility (log S) based on two machine-learning models and AlogPS 2.1.[29] Among the numerous solubility prediction models proposed to date, AlogPS 2.1 is known for its high accuracy, reading it reliable.[30] Table 2 reveals that the accuracies of our models of our models derived from MAPs and IAPs are comparable with that of AlogPS 2.1 signifying the successful development of effective models.

Table 2: Metrics for AlogPS 2.1 and the proposed GNNs models with MAPs and IAPs

	RMSE	R^2
MAPs	0.60±0.02	0.94
IAPs	0.53±0.02	0.95
AlogPS 2.1	0.57	0.93

In chemistry, attributions are predominantly discussed based on atoms. In this study, we employed modified IGs to obtain the attribution for each atom. Figure 1 displays colormaps of the IGs used for methanol l (MeOH) and 2,2',3,3',4,4',5,5'-Octachlorobiphenyl (PCB194). Our choice of these molecules was based on the observation that MeOH was the most soluble while PCB194 was the least soluble (insoluble) among the molecules in the test dataset. In the figure, atoms contributing more to solubility appear in red (> 0), whereas those contributing more to insolubility appear in blue (< 0). For methanol, irrespective of the explanatory variables, the portion corresponding to the hydroxyl group consistently appears in red, indicating its successful representation as a hydrophilic moiety. This observation is consistent, making it a valuable indicator for attribution.

As stated in the Introduction section, the attribution assignment of atom groups has been

extensively researched in the field of chemistry. Among the developed methods, the Hansch–Fujita method is the most prominent attribution assignment approach. The researchers demonstrated that the changes in the hydrophobicity levels of eight parent structures induced by the introduction of substituents could be defined as the substituent constant (π), which could be leveraged to discuss the significance of each functional group in structure-activity studies. Thus, this substituent constant π represents an attribution assignment of atom groups based on an experimental approach. In this study, we adopted the π value of phenoxyacetic acids as they have a comprehensive dataset comprising the eight parent structures. Subsequently, we investigated the correlation between the IGs and π values for 16 functional groups (Scheme 1), using molecules that were not double-counted (excluding functional groups such as sulfonic and phosphate groups, which are not included in the substituent constant π).

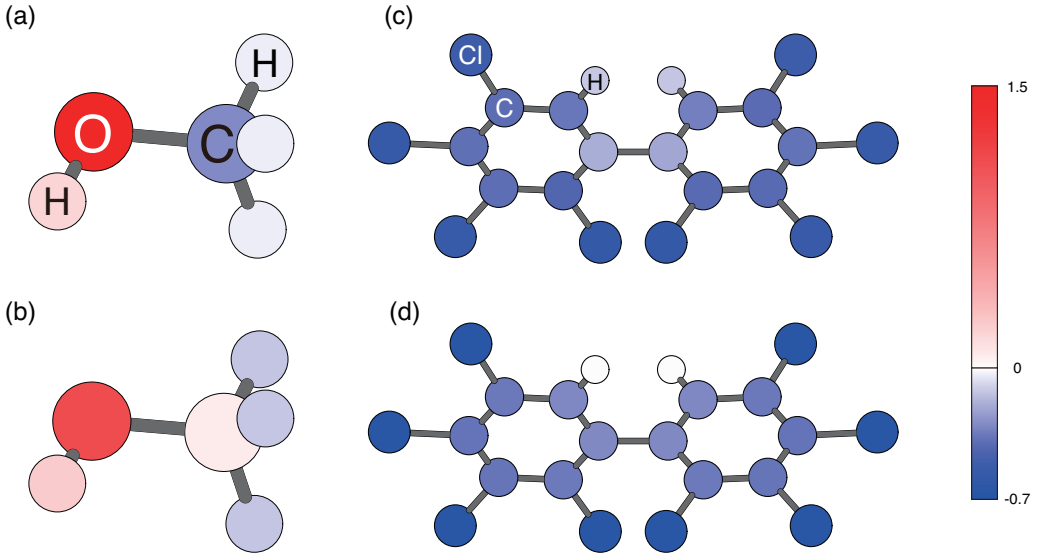
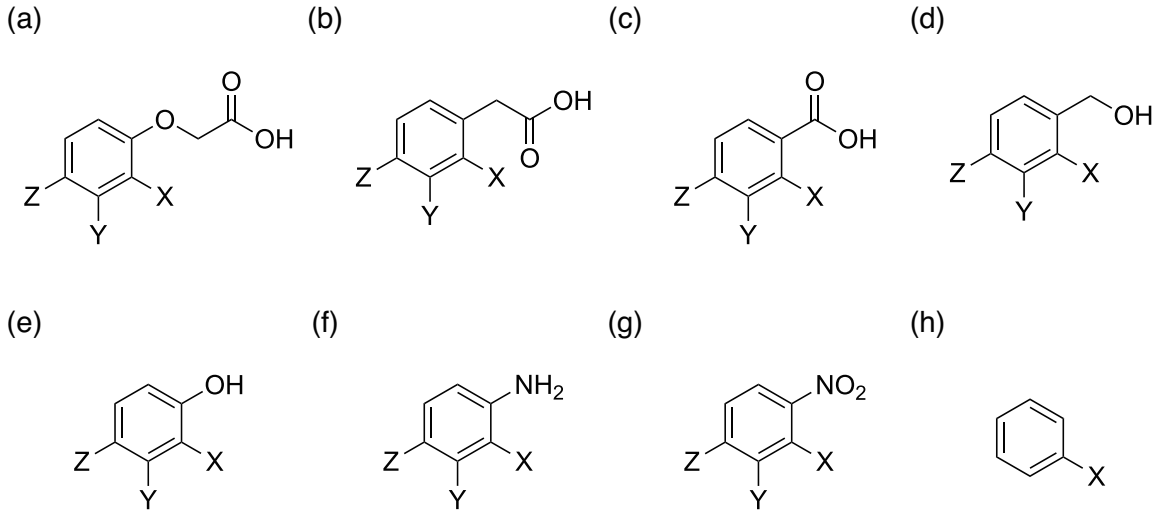


Figure 1: Colormaps of the IGs of methanol ((a) and (b)) and 2,2',3,3',4,4',5,5'-Octachlorobiphenyl ((c) and (d)). The IGs were computed considering the MAPs ((a) and (d)) and IAPs ((b) and (d)).

The IGs and π values were determined based on log S and log P, respectively. Therefore, when a functional group was deemed important for water solubility, the IGs values became significantly positive, whereas the π value became significantly negative. For simplicity, we

changed the signs of the IGs values in Figure 2. As depicted in Figure 2, the π and IGs values exhibit good correlations, with minimal dependence on the descriptor. Although the approaches adopted to derive attributions largely differ between the Hansch–Fujita method and our method, the obtained attributions exhibit a similar characteristic.



Scheme 1: substituent constant(π) defined based on the eight different parent structures

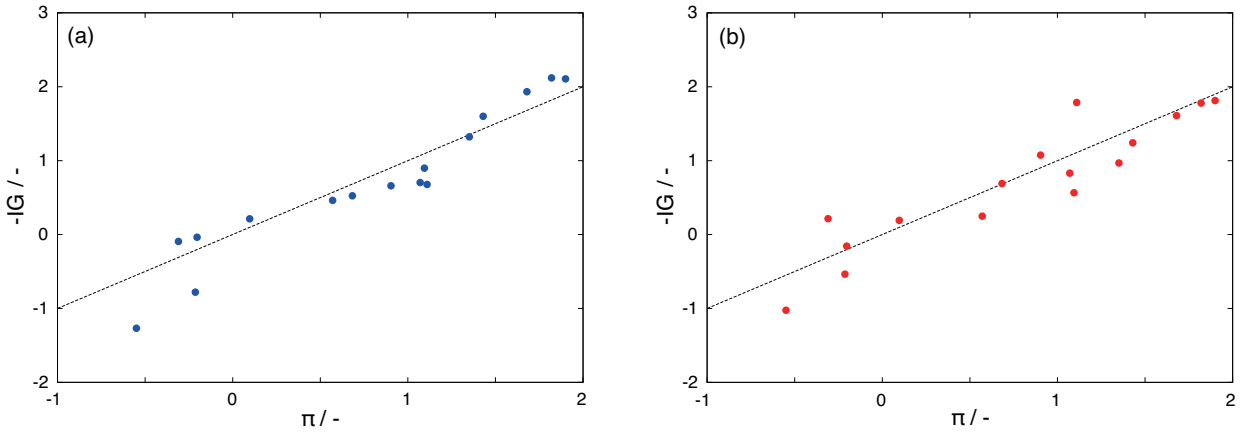


Figure 2: Correlations between IGs and the substituent constant (π) computed using (a) MAPs and (b) IAPs

While Figure 2 indicates a good correlation between the π and IGs values, certain substituents exhibit significant discrepancies. In particular, the disparity for the OH group computed using MAPs is large compared to those for the other groups. To understand the reason for this, we must consider structural dependency in attribution assignment. For instance, in

the Hansch–Fujita method, the π value of the OH group is determined using 15 datapoints based on the following parent compounds: *ortho*-, *meta*-, and *para*-isomers (Scheme 1). In this study, we plot the variability of each attribution for two types of substituents (Figure 3) to elucidate the structural dependency in attribution assignment.

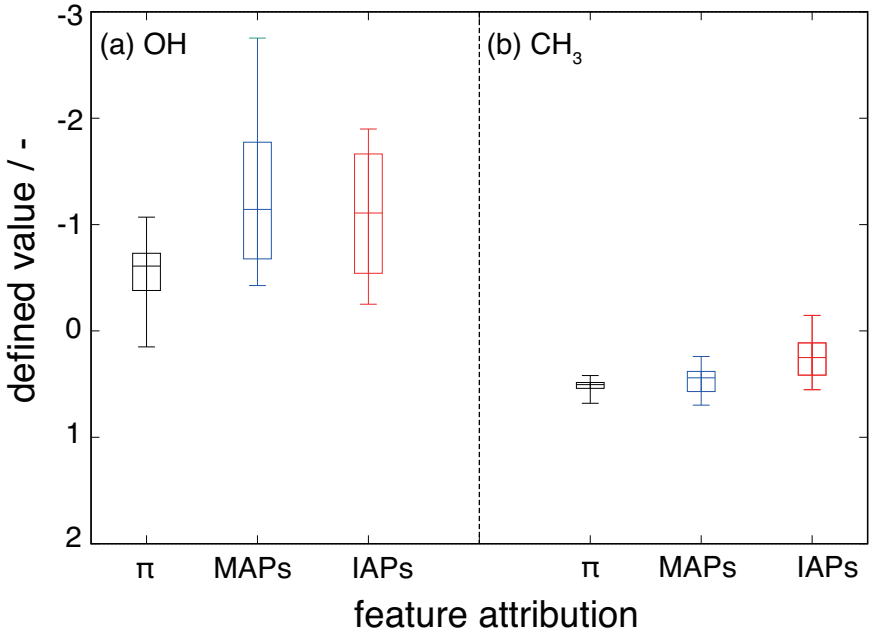


Figure 3: Box-and-whisker plot depicting the degree of variation of each attribution for substituents OH (a) and CH_3 (b). Here, the attributions represent the substituent constant (π), MAPs, and IAPs.

As depicted in Figure 3 (a), the IGs and π values of the OH group exhibit the largest difference (Figure 2). Conversely, the CH_3 group, with a larger dataset, displays relatively minor differences, implying that the contribution of the OH group to water solubility is considerably affected by its molecular structure. Figure 3 illustrates that the notable discrepancy in the IGs and π values of the OH group displayed in Figure 2 does not stem from numerical errors but provides valuable insights into the influence of molecular structures on water solubility.

Conclusion

This study examined and contrasted the feature attribution concepts of machine learning and chemistry. Both learning models, developed using water solubility as a predictor, were found to exhibit good prediction accuracies. Moreover, the IGs results indicated similar feature attributions, implying the development of good-quality learning models. The developed feature attributions exhibited good correlations with the feature attributions based on Hansch's π from the field of chemistry. While some functional groups presented large discrepancies between the machine-learning and chemistry-based feature attributions, we demonstrated that these discrepancies were attributable to the influence of molecular structures. Through this comparison, we can conclude that our machine-learning models predict water solubility using feature attributions similar to those obtained by chemists.

Acknowledgement

This study was supported by JST, PRESTO Grant Number JPMJPR21C9.

Author Contributions

S.T. and K.N. analyzed and interpreted the results, and S.T. drafted the manuscript. S.T. and K.N. contributed equally to this work. D.Y. and K.S. prepared molecular–atomic properties (MAPs) using quantum mechanical calculations. D.Y. was involved in planning and supervised the work. All authors discussed the results and commented on the manuscript.

References

(1) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.

(2) Dong, S.; Wang, P.; Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **2021**, *40*, 100379.

(3) Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115.

(4) Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160.

(5) Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C. J.; Müller, K.-R. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proceedings of the IEEE* **2021**, *109*, 247–278.

(6) Kikutsuji, T.; Mori, Y.; Okazaki, K.-i.; Mori, T.; Kim, K.; Matubayasi, N. Explaining reaction coordinates of alanine dipeptide isomerization obtained from deep neural networks using Explainable Artificial Intelligence (XAI). *J. Chem. Phys.* **2022**, *156*, 154108.

(7) Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *CoRR* **2016**, *abs/1602.04938*.

(8) Lundberg, S. M.; Lee, S. A unified approach to interpreting model predictions. *CoRR* **2017**, *abs/1705.07874*.

(9) Ishida, S.; Terayama, K.; Kojima, R.; Takasu, K.; Okuno, Y. Prediction and Interpretable Visualization of Retrosynthetic Reactions Using Graph Convolutional Networks. *J. Chem. Inf. Model.* **2019**, *59*, 5026–5033.

(10) Wöhler, F.; Liebig, J. Untersuchungen über das Radikal der Benzoesäure. *Ann. Phar.* **1832**, *3*, 249–282.

(11) Liebig, J. Ueber die Constitution des Aethers und seiner Verbindungen. *Ann. Phar.* **1834**, *9*, 1–39.

(12) Dumas, J.; von Liebig, J. Note on the present state of organic chemistry. *CR Hebd. Seances Acad. Sci.* **1837**, *5*, 567–572.

(13) Dumas, J. Mémoire sur la constitution de quelques corps organiques et sur la théorie des substitutions. *CR Hebd. Seances Acad. Sci.* **1839**, *8*, 609–633.

(14) Williamson, A. XLV. Theory of ætherification. *London Edinburgh Philos. Mag. & J. Sci.* **1850**, *37*, 350–356.

(15) Hofmann, A. W. XV.—Researches on the volatile bases. *Q. J. Chem. Soc.* **1849**, *1*, 159–173.

(16) Gerhardt, C. Recherches sur les acides organiques anhydres. *Ann. Chim. Phys.* **1853**, *37*, 285–342.

(17) Constable, E. C.; Housecroft, C. E. Before Radicals Were Free – the Radical Particulier of de Morveau. *Chemistry* **2020**, *2*, 293–304.

(18) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

(19) Li, X.; Gao, X.; Shi, W.; Ma, H. Design Strategies for Water-Soluble Small Molecular Chromogenic and Fluorogenic Probes. *Chem. Rev.* **2014**, *114*, 590–659.

(20) Zhu, C.; Liu, L.; Yang, Q.; Lv, F.; Wang, S. Water-Soluble Conjugated Polymers for Imaging, Diagnosis, and Therapy. *Chem. Rev.* **2012**, *112*, 4687–4735.

(21) Wang, L. G.; Montaño, A. R.; Combs, J. R.; McMahon, N. P.; Solanki, A.; Gomes, M. M.; Tao, K.; Bisson, W. H.; Szafran, D. A.; Samkoe, K. S.; Tichauer, K. M.; Gibbs, S. L. OregonFluor enables quantitative intracellular paired agent imaging to assess drug target availability in live cells and tissues. *Nat. Chem.* **2023**, *15*, 729–739.

(22) Aktay, G.; Du, Y.-Z.; Torrado, J.; Savjani, K. T.; Gajjar, A. K.; Savjani, J. K. Drug Solubility: Importance and Enhancement Techniques. *ISRN Pharm.* **2012**, *2012*, 195727.

(23) Barrett, J. A.; Yang, W.; Skolnik, S. M.; Belliveau, L. M.; Patros, K. M. Discovery solubility measurement and assessment of small molecules with drug development in mind. *Drug Discov. Today* **2022**, *27*, 1315–1325.

(24) Liu, X.; Zhao, L.; Wu, B.; Chen, F. Improving solubility of poorly water-soluble drugs by protein-based strategy: A review. *Int. J. Pharm.* **2023**, *634*, 122704.

(25) Yokogawa, D.; Suda, K. Interpretable Attribution Assignment for Octanol–Water Partition Coefficient. *J. Phys. Chem. B* **2023**, *127*, 7004–7010.

(26) Yokogawa, D.; Suda, K. Feature selection in molecular graph neuralnetworks based on quantum chemical approaches. *Digital Discovery* **2023**, *2*, 1089–1097.

(27) Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z. *Deep Learning for the Life Sciences*; O'Reilly Media, 2019.

(28) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, π, Derived from Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.

(29) Tetko, I. V.; Bruneau, P. Application of ALOGPS to predict 1-octanol/water distribution coefficients, logP, and logD, of AstraZeneca in-house database. *J. Pharm. Sci.* **2004**, *93*, 3103–3110.

(30) Kenney, D. H.; Paffenroth, R. C.; Timko, M. T.; Teixeira, A. R. Dimensionally reduced machine learning model for predicting single component octanol-water partition coefficients. *J. Cheminform* **2023**, *15*, 9.