A Machine Learning Approach for Predicting Defluorination of Per- and Polyfluoroalkyl Substances (PFAS) for Their Efficient Treatment and Removal

Akber Raza,[a] Lihua Xu,[b] Sharma S.R.K.C. Yamijala,[b,c] Chao Lian,[b,c,d] Hyuna Kwon,[b] and Bryan M. Wong[a,b,c,d]*

[a]Department of Electrical & Computer Engineering, [b]Department of Chemical & Environmental Engineering, [c]Materials Science & Engineering Program, and [d]Department of Physics & Astronomy

University of California, Riverside, Riverside, CA 92521, United States

*Corresponding author. E-mail: bryan.wong@ucr.edu; Web: http://www.bmwong-group.com

Abstract

We present the first application of machine learning on per- and polyfluoroalkyl substances (PFAS) for predicting and rationalizing carbon-fluorine (C–F) bond dissociation energies to aid in their efficient treatment and removal. Using a variety of machine learning algorithms (including Random Forest, Least Absolute Shrinkage and Selection Operator Regression, and Feed-forward Neural Networks), we were able to obtain extremely accurate predictions for C–F bond dissociation energies (with deviations less than 0.70 kcal/mol) that are *within chemical accuracy* of the PFAS reference data. In addition, we show that our

machine learning approach is extremely efficient (requiring less than 10 minutes to train the data and less than a second to predict the C–F bond dissociation energy of a new compound) and only needs knowledge of the simple chemical connectivity in a PFAS structure to yield reliable results – without recourse to a computationally expensive quantum mechanical calculation or a three-dimensional structure. Finally, we present an unsupervised machine learning algorithm that can automatically classify and rationalize chemical trends in PFAS structures that would otherwise have been difficult to humanly visualize/process manually. Collectively, these studies (1) comprise the first application of machine learning techniques for PFAS structures to predict/rationalize C–F bond dissociation energies and (2) show immense promise for assisting experimentalists in the *targeted* defluorination of specific bonds in PFAS structures (or other unknown environmental contaminants) of increasing complexity.

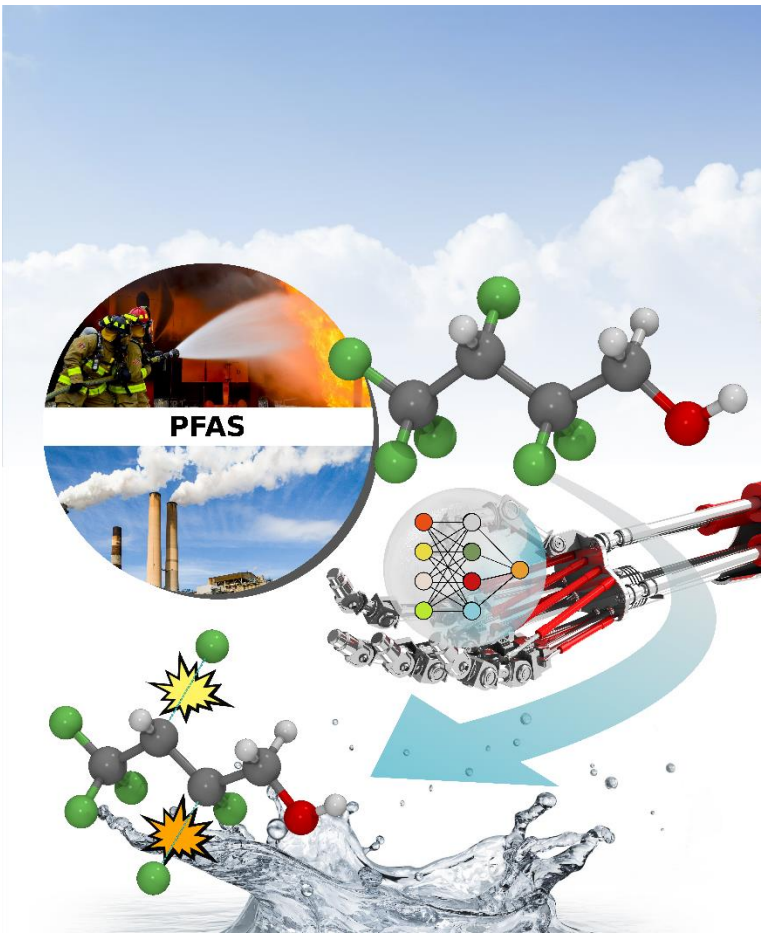


Table of Contents Figure

Introduction

The efficient treatment and removal of per- and polyfluoroalkyl substances (PFAS) continues to garner immense interest due to their deleterious health effects and their widespread presence in surface and groundwater sources. In particular, PFAS have been detected in hundreds of locations worldwide and have been linked to harmful health effects in the liver, kidneys, blood, and immune system.[1, 2] One of the most common ways that PFAS contaminants impact the environment is through their presence in high-performance firefighting foams, which are used to extinguish fuel-based fires.[3-5] In addition, these compounds are also used in many non-stick and stain-repellent household products and can be unintentionally introduced into the environment when manufacturing waste is improperly disposed of.[6, 7] Because PFAS molecules contain a variety of strong carbon-fluorine (C–F) bonds, they become persistent in the environment and are extremely difficult to treat/remove once they have contaminated water resources.[8, 9]

While long-chain PFAS molecules such as perfluorooctanoic acid (PFOA) and perfluorooctanesulfonic acid (PFOS) are no longer manufactured in the US, many smaller-chain PFAS are still in widespread use. For example, aqueous film-forming foams used in high-performance fire extinguishers contain at least hundreds of PFAS structures,[3, 4, 10, 11] many of which are composed of fluorocarbon chains of varying lengths and a diversity of organic head groups. Because of this immense chemical diversity, prior work by us[12] and others[13-15] have shown that the efficient cleanup of these contaminated sites can only be attained if *all* the PFAS species (not just PFOA or PFOS alone) are properly accounted for and treated. However, the sheer number and variety of PFAS contaminants is immense, and an in-depth experimental (or even theoretical) study of this nearly limitless number of chemical compounds is extremely difficult.

We show for the first time that PFAS species and their defluorination mechanisms are natural candidates for harnessing machine learning approaches. Over the past few years, our society has witnessed an unprecedented growth in the use of machine learning and artificial intelligence in technological applications such as automated medical diagnostics software,[16] handwriting recognition,[17] computer

vision,[18] and autonomous vehicles[19] (to name just a select few). While there has been prior work on using machine learning approaches for chemical/material applications,[20-23] these advanced algorithms have not been previously applied to PFAS-based environmental studies with the specific purpose of predicting/rationalizing C–F bond dissociation energies to aid in their efficient treatment/removal. In this work, we show that the sheer variety of PFAS structures and defluorination mechanisms naturally lends itself to a data-driven machine learning approach, which can subsequently yield both efficient and accurate results. By training a computer on over 560 distinct C–F bond energies found in representative PFAS molecules (cf. **Figure 1**), we show that a machine learning algorithm can give extremely accurate predictions (deviations less than 0.70 kcal/mol) that are *within chemical accuracy* of the PFAS reference data. Furthermore, our machine learning approach is extremely efficient and only requires knowledge of the chemical connectivity of a PFAS molecule (i.e., a user can simply sketch an arbitrary PFAS structure by hand on paper, which can be entered into our machine-learned model to predict an accurate C–F bond dissociation in a few seconds). Finally, we show that advanced machine learning algorithms can be used to (1) automatically classify the C–F bonds across a variety of PFAS structures and (2) rationalize chemical trends in these structures without any human intervention. Collectively, these studies comprise the first application of machine learning techniques for PFAS structures, with the specific purpose of predicting/rationalizing C–F bond dissociation energies to aid in their efficient treatment and removal.

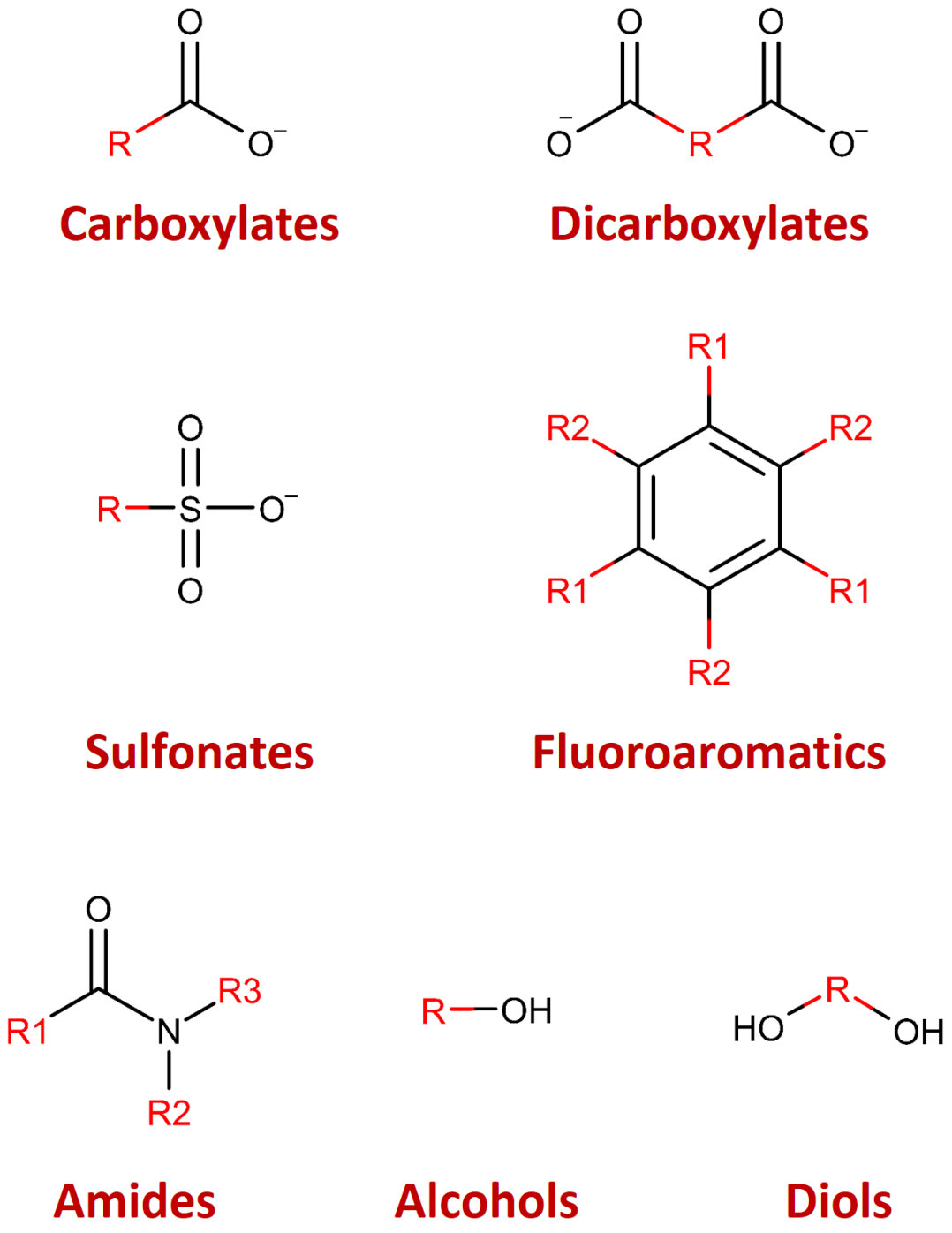


Figure 1. Chemical families of all the PFAS molecules examined in this work. The red-colored R groups in each structure represent various C–F-containing molecular units (564 distinct C–F bonds in total – see Supporting Information for a complete set of structures) that encompass a broad range of molecular chains, branched structures, and other chemical functional groups.

Materials and Methods

To predict and understand C–F bond dissociation energies in various PFAS structures of environmental importance, we utilized a variety of machine learning techniques that include Random

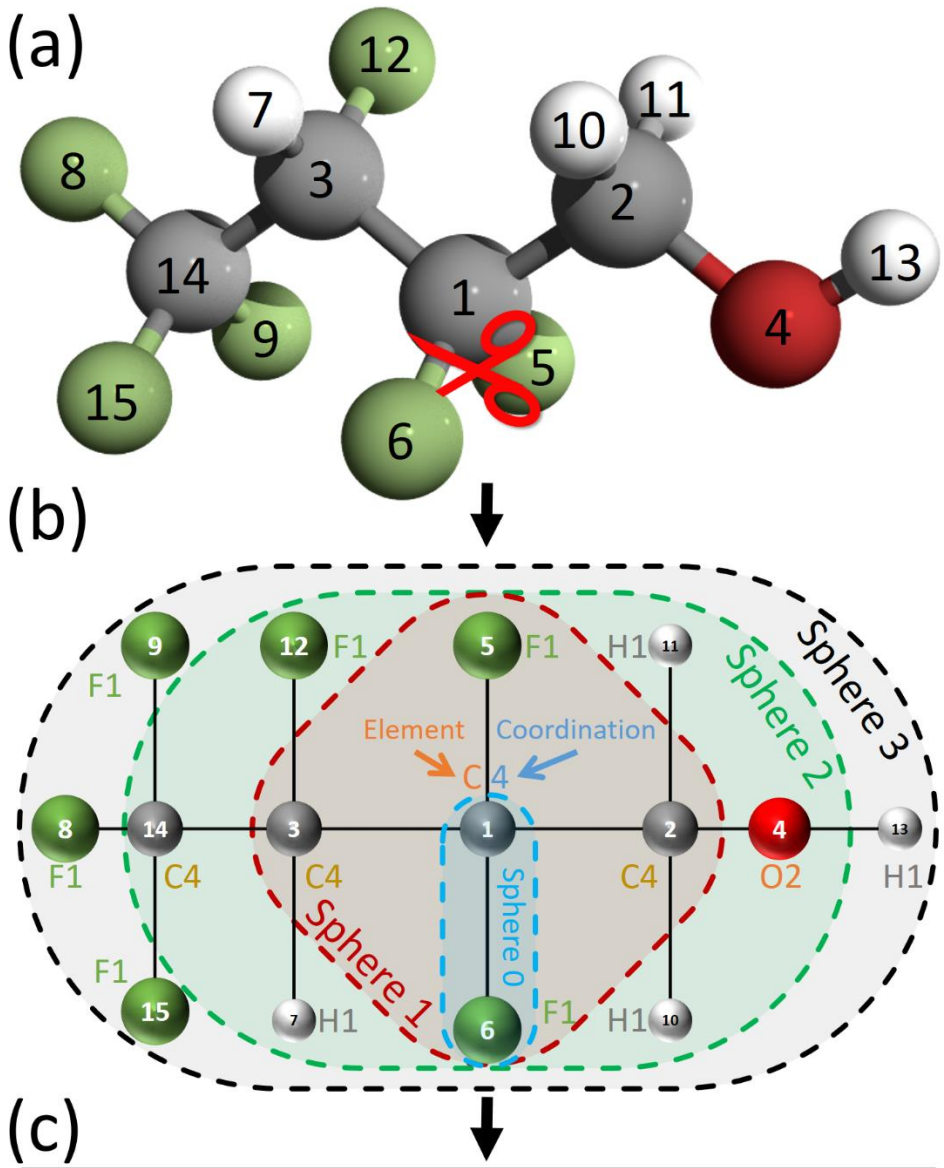
Forest, Least Absolute Shrinkage and Selection Operator (LASSO) Regression, Feed-forward Neural Network (FNN), and t-distributed Stochastic Neighbor Embedding (t-SNE) algorithms.[17] The first three algorithms are categorized as supervised machine learning techniques since they require training data that is *pre-labeled* by a teacher/expert (i.e., in a "supervised" fashion). Regression methods in each of these algorithms are then subsequently used to find patterns in the data via different optimization techniques. In the context of the PFAS structures examined in this work, the "labels" are the DFT-computed C–F bond dissociation energies, and the desired output from each of these algorithms is an accurate prediction of a C–F bond dissociation energy of a PFAS structure *not in the original training data set*. The last t-SNE algorithm mentioned previously is an unsupervised machine learning technique since no training labels are given, and the algorithm must find patterns in the data on its own (i.e. in an "unsupervised" fashion) to determine how to categorize/cluster the provided data. In the context of the present study, the goal of the t-SNE machine learning algorithm is to categorize all of the PFAS C–F bond dissociation energies into clusters/families (described further in the Results and Discussion section) to understand which chemical functional groups are responsible for the observed bond dissociation energies. Further algorithmic details for each of the four machine learning approaches used in this work are provided in the Supporting Information.

In any machine learning problem, the data plays a critical role; as such, all of the machine learning algorithms in this work utilized C–F bond dissociation energies that were computed at the dispersion-corrected[24-26] B3LYP+D3BJ/6-311+G(2d,2p) level of theory in conjunction with Truhlar's SMD solvation model to implicitly simulate an aqueous environment.[27] We have chosen to use this particular level of theory since the B3LYP+D3BJ functional (1) exhibits the mathematically exact R^{-6} asymptotic dependence required to accurately describe dispersion effects in intermolecular/dissociation interactions and (2) provides accurate predictions in close agreement with experimental results, particularly for PFAS-based compounds and their hydrogen-bonded complexes.[28-34] Harmonic frequency calculations for all of the 564 C–F bond-dissociation reactions (the complete data set for all 564 structures needed to run our machine learning algorithms can be downloaded as an external .zip file in the Supporting Information), and the bond

dissociation energies were calculated from the total enthalpies (which include thermal corrections from translational, rotational, and vibrational degrees of freedom) for each of the chemical species. Further computational details of these DFT calculations are given in prior work.[12, 35]

To utilize the various machine learning approaches, we require a set of chemical descriptors to enable a computer to autonomously and rapidly process this extensive data set. In order to be practical/efficient, the chosen descriptors should not be expensive to compute and, therefore, should satisfy the following 4 requirements for describing PFAS molecular structures: the desired chemical descriptor should (1) use a simple algorithm, (2) not rely on a quantum chemistry calculation to be carried out, (3) not require an optimized 3D geometry of the molecule, and (4) not explicitly use bond orders (i.e., single or double bonds). These requirements for descriptors were specifically chosen since any descriptor requiring a complex calculation (such as an optimized 3D geometry, for example) would defeat the purpose of using machine learning at all, because creating the descriptor, in essence, would just be as complicated as doing a quantum calculation/optimization itself. To satisfy these requirements, we have chosen to utilize the chemically intuitive bond descriptor scheme by Qu et al.,[36] which utilizes the Chemistry Development Kit (CDK) libraries and is available as an open-source code written in the Java programming language.[37] Although the prior work by Qu et al. originally focused on organic molecules containing H, C, N, O, and S atoms, we have modified their Java-based source code (available for download at http://www.bmwong-group.com/software) to include various C–F bond descriptors for the PFAS structures examined in this study. Within this scheme, geometric boundaries or "spheres" are used to encode the distance between the target bond of interest and the other atoms in the molecule. To visually conceptualize this more easily, **Figure 2** gives a graphical representation of the various spheres and bond descriptors for dissociating a specific C–F bond in a prototypical PFAS molecule. In this scheme, each atom is labeled according to its element name and coordination number. For example, the carbon atom in a methane molecule would be labeled as C4 (since it has a coordination of 4 and is bonded to 4 hydrogen atoms), the carbon atom in ethylene would be labeled as C3, and the carbon atom in acetylene would be given a C2 label. For all of the molecules in this study, we consider only 15 atom types: C2, C3, C4, H1, N1, N2, N3, N4, O1, O2, S1,

S2, S3, S4, and F1. As shown in **Figure 2**, the classification of atoms in different spheres is based on their distance to the target bond (this distance is defined in Ref. 36 as the number of covalent bonds [both single and double bonds only count as 1 bond] between that atom and an atom in the target bond on the shortest possible path). For example, atoms 1 and 6 are involved in the target C–F bond to be dissociated and belong to sphere 0; atoms 2, 3, and 5 are one bond away from the target 1–6 bond and, therefore, belong to sphere 1; atoms 4, 7, 10, 11, 12, and 14 are two bonds away from the target bond and belong to sphere 2, and so on.



	C2	C3	C4	H1	N1	N2	N3	N4	O1	O2	S1	S2	S3	S4	F1
Sphere 0			1												1
Sphere 1			2												1
Sphere 2			1	3						1					1
Sphere 3				1											3

Figure 2. Schematic of the chemical descriptors and data used by our machine learning algorithms to autonomously and rapidly process C–F bond dissociation energies in general PFAS structures. Panel (a) depicts a specific target C–F bond to be dissociated, and panel (b) shows the atom labeling scheme and spheres used to construct the point descriptor table in panel (c). It is important to note that the 3D and 2D structures in panels (a) and (b) are shown only for illustrative purposes and are *not used* by our machine learning algorithms, whereas the simple pair descriptor table in panel (c) encodes the actual data that is used by the machine learning algorithms in this work.

With the atom labels and spheres properly defined for each unique PFAS C–F bond, this information can be encoded as a point descriptor table, which is shown in **Figure 2c**. Simply put, the individual entries in this table correspond to the number of atom types in a specific sphere. Next, a pair descriptor table (not shown in **Figure 2**) can be constructed, which contains information on the number of pairs of atom types in specific spheres at designated distances between them. It should be mentioned that there are other variants/sub-categories of point and pair descriptors (such as element point descriptors, connection number point descriptors, bond-breaking difference pair descriptors, and molecular descriptors), which are discussed in detail in Ref. 36. Finally, the collected data was further cleaned by removing duplicate C–F bonds that contained the same bond dissociation energy and bond descriptors. Together, the simple point and pair descriptors (which, again, do not contain any explicit 3D geometrical/quantum-mechanical data and also satisfy the 4 requirements mentioned previously for efficiently describing PFAS molecular structures) can be automatically processed via machine learning algorithms to accurately predict a C–F bond dissociation energy of a general PFAS molecule.

Results and Discussion

In this section, we compare the performance of each of the machine learning algorithms on predicting C–F bond dissociation energies for a variety of PFAS molecular structures. Among the three supervised machine learning approaches examined in this work (i.e., Random Forest, LASSO Regression, and FNN), the FNN algorithm (available as a standalone Python code within a .zip file in the Supporting Information), is the most complex and was originally designed to "learn" highly convoluted patterns in the underlying data, as discussed below. **Figure 3** shows a schematic of the one-hidden layer neural network utilized in our study, which consists of one input layer, one fully connected hidden layer, and one output neuron. In the context of our current study, the various PFAS bond descriptors were used in the input layer, and the output was the predicted PFAS bond dissociation energy. Within the hidden layer, a numerical mapping occurs between the input and output layers, which is described/controlled by mathematical activation functions. For our study, we utilized the rectified linear unit (ReLU) activation function,[38] to

allow greater flexibility in learning the mapping between the input bond descriptors and the outputted bond dissociation energy.

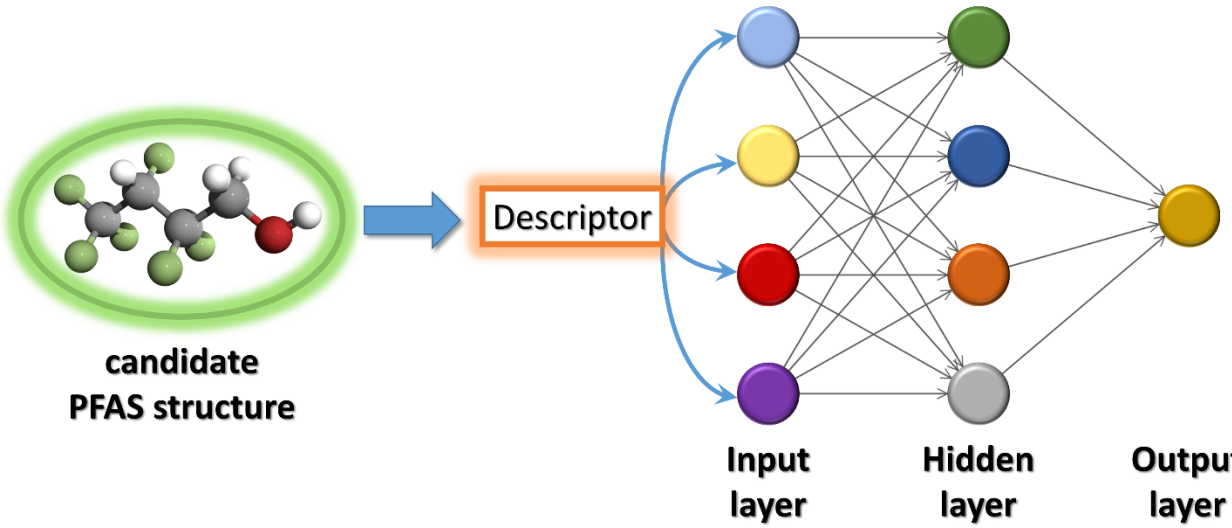


Figure 3. Schematic of the Feed-forward Neural Network used to predict PFAS bond dissociation energies in this work. PFAS bond descriptors are used in the input layer, which is subsequently processed by a hidden layer to learn a (non-linear) mathematical mapping for generating predicted PFAS bond dissociation energies at the output layer.

As is customary for any supervised machine learning approach, we subdivided the data into two sets: a training and test set, described further below. Among the 564 unique C–F bond dissociation energies, 414 randomly chosen C–F bonds were utilized in the training set to enable each of the three supervised machine learning algorithms to autonomously learn patterns in the data. The remaining 150 C–F bond dissociation energies were then subsequently utilized to test the predictive performance of each machine learning algorithm, as shown in **Figure 4**. The diagonal line in all of these figures represents an ideal 100% agreement between that particular machine learning algorithm and the corresponding reference DFT bond dissociation energies. Among the three supervised machine learning algorithms, **Figure 4** shows that the FNN approach yields predictions that are in excellent agreement with the reference DFT data with an impressive R^2 value of 0.93. To make an additional comparison across the various supervised machine learning algorithms, we calculated the Mean Absolute Deviation (MAD) and Root Mean Squared Error (RMSE) from our reference DFT bond dissociation energies for the Random Forest, LASSO Regression, and FNN approaches in **Table 1** based on the 4-sphere bond descriptor model described previously (it is

worth mentioning that we also calculated MADs and RSMEs for each of the machine learning models as a function of the number of sphere descriptors (i.e., 3, 4, and 7) and found that the 4-sphere bond descriptor model gave the best results, which is consistent with the previous study on organic molecules by Qu et al.[36]). The most salient result of **Figure 4** and **Table 1** is that the FNN machine learning algorithm gives strikingly accurate predictions (MAD = 0.70 kcal/mol) that are *within chemical accuracy* of the PFAS bond dissociation reference data. In other words, the error introduced by the FNN algorithm is quite small and within the accuracy of the employed theoretical method (namely B3LYP+D3BJ/6-311+G(2d,2p)). It is important to note that no high-level wavefunction-based benchmark datasets on PFAS molecules currently exist (to the best of our knowledge) to judge the accuracy of the B3LYP+D3BJ functional; however, based on various benchmark calculations on other datasets,[39-43] the accuracy of this functional is within 0.2-2.5 kcal/mol.

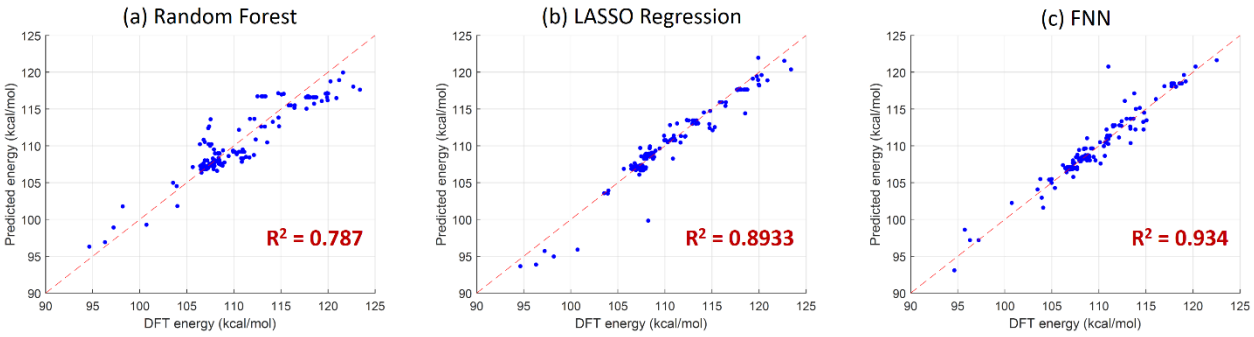


Figure 4. Comparison of the accuracy among the (a) Random Forest, (b) LASSO Regression, and (c) FNN machine learning algorithms for predicting C–F bond dissociation energies in the PFAS test set. The diagonal line in each figure represents a perfect match between that particular machine learning algorithm and the corresponding reference DFT data.

Table 1: Mean Absolute Deviation (MAD) and Root Mean Squared Error (RMSE) Values for Each of the Three Supervised Machine Learning Algorithms on Predicting PFAS Bond Dissociation Energies. The Values in Parentheses Correspond to MADs and RSMEs Computed on a $\log_{10}(K)$ Scale (i.e., MAD/RSME Values Divided by $2.303 \cdot R \cdot T$).

	Random Forest	**LASSO Regression**	**Feed-forward Neural Network (FNN)**
MAD (kcal/mol)	2.42 (1.77)	1.96 (1.44)	0.70 (0.51)
RMSE (kcal/mol)	2.65 (1.94)	1.87 (1.37)	1.22 (0.89)

As a final example of utilizing machine learning approaches to understand PFAS defluorination mechanisms, **Figure 5** plots the classifications/trends resulting from the t-SNE algorithm, which is an unsupervised machine learning approach for finding patterns in the data without any prior training labels from the user. In simple terms, the t-SNE algorithm allows the visualization of high-dimensional data as two-dimensional "clusters" where data points grouped within a cluster share similar characteristics with each other. In the context of our study on PFAS molecules, **Figure 5** shows that C–F bonds with similar bond dissociate energies are *automatically clustered* together by the t-SNE algorithm. Specifically, C–F bonds with large dissociation energies (enclosed by the yellow ellipse for clarity) are mostly attached to the terminal end groups of molecules. In contrast, C–F bonds with low dissociation energies (enclosed by the blue ellipse) arise from branched molecules. Finally, clusters enclosed by the lime green ellipse at the bottom of **Figure 5** represent C–F bonds that are mostly adjacent to the terminal carbon of the molecule, whereas clusters enclosed by the teal green ellipse arise from C–F bonds that exist in the middle of the molecular chain. It is important to re-iterate that these clusters/classifications were *automatically* chosen by the unsupervised t-SNE machine learning algorithm (without human intervention). As such, these results demonstrate that the t-SNE algorithm can be a useful tool for automatically classifying and rationalizing chemical trends in PFAS structures that would otherwise have been difficult to humanly visualize/process manually.

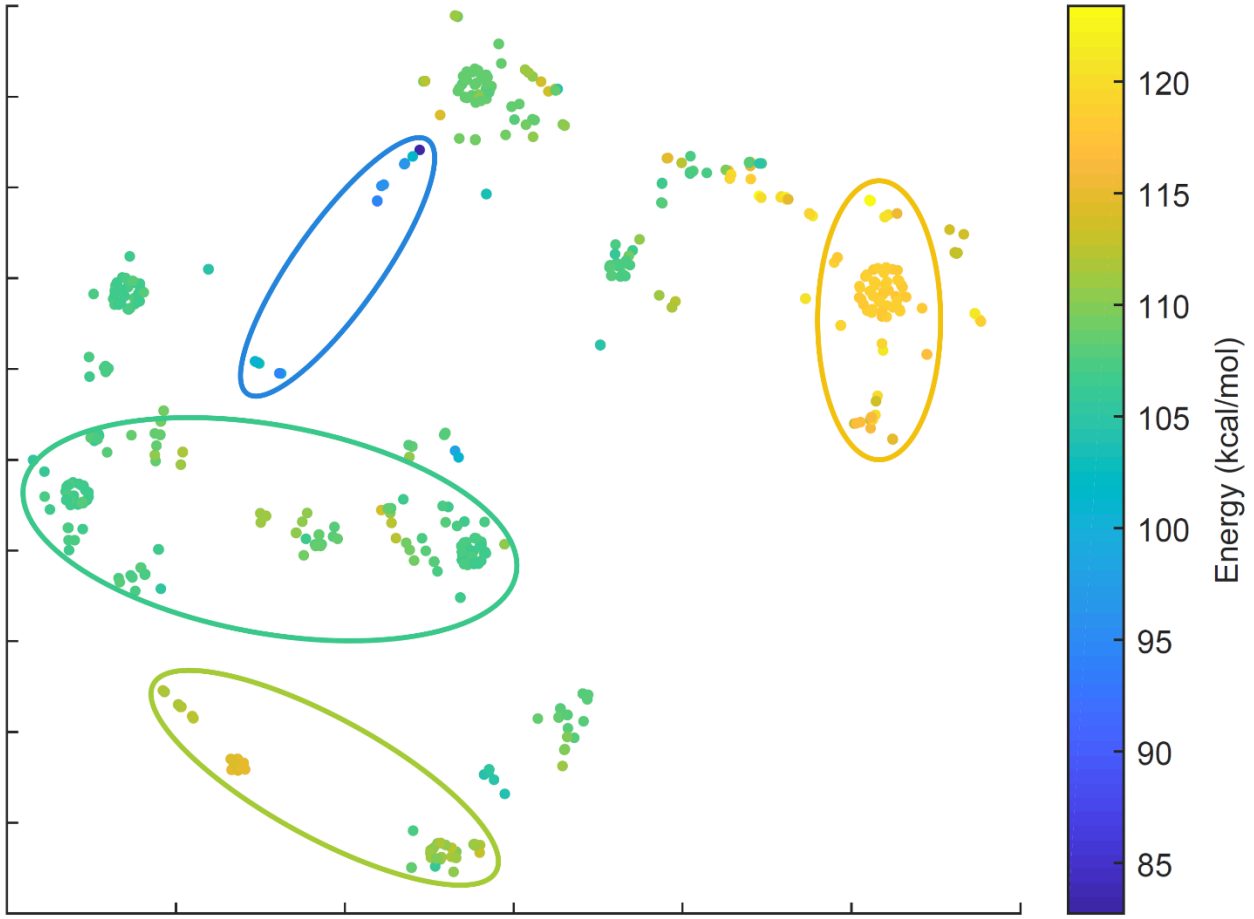


Figure 5. Clustering/categorization of C–F bond dissociation energies in various PFAS structures automatically predicted by the t-SNE machine learning algorithm. The yellow ellipse encloses C–F bonds attached to the terminal end groups of molecules and have large dissociation energies; in contrast, the blue ellipse contains C–F bonds in branched molecules and have low dissociation energies. Clusters enclosed by the lime green ellipse are associated with C–F bonds adjacent to the terminal carbon of the molecule, whereas clusters enclosed by the teal green ellipse arise from C–F bonds in the middle of the molecular chain.

In conclusion, we have presented the first application of machine learning on PFAS structures with the specific purpose of predicting/rationalizing C–F bond dissociation energies to aid in efficient PFAS treatment and removal. Using a variety of supervised machine learning algorithms (including Random Forest, LASSO Regression, and FNN approaches), we were able to demonstrate that the sheer variety of PFAS structures and defluorination mechanisms naturally lends itself to a machine learning approach, which can subsequently yield both efficient and accurate results. In terms of efficiency, we have shown that our machine-learned model only requires knowledge of the simple chemical connectivity in a PFAS structure (i.e., neither a 3D geometry nor even a rough estimate of bond lengths/orientations are required)

to yield reliable results. In other words, a user can simply sketch an arbitrary PFAS structure by hand on paper that can be entered into our machine-learned model to obtain accurate dissociation energies for any of its C–F bonds in less than a second (without having to do a computationally expensive quantum mechanical optimization or frequency calculation). In terms of accuracy, we have shown that the FNN machine learning algorithm gives extremely accurate predictions (deviations less than 0.70 kcal/mol) that are *within chemical accuracy* of the PFAS reference data; therefore, this approach can be used to predict bond dissociation energies for other unexplored PFAS molecules of environmental significance. In our final application of machine learning, we have shown that the unsupervised t-SNE machine learning algorithm can automatically categorize and rationalize chemical trends in PFAS structures that would otherwise have been difficult to humanly visualize/process manually. Taking these results together, the combined use of the t-SNE algorithm with the highly accurate/efficient FNN machine learning approach shows immense promise for assisting experimentalists in the *targeted* defluorination of specific bonds in PFAS structures (or other unknown environmental contaminants) of increasing complexity.

Supporting Information

Complete data set for all 564 structures (in CSV format within a compressed .zip file) required as input for most machine learning algorithms, standalone Python code for the FNN algorithm in the same external .zip file, additional algorithmic details for each of the four machine learning approaches used in this work, additional details on the accuracy of the supervised machine learning algorithms, and additional details on the choice of the training and test dataset. This information is available free of charge on the ACS Publications website.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CHE-1808242. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Comet

computing cluster at the University of California, San Diego through allocation TG-ENG160024. Prof. Jinyong Liu and Ms. Sharmistha Bardhan are acknowledged for several helpful discussions.

References

1. Butenhoff, J. L.; Rodricks, J. V., Human Health Risk Assessment of Perfluoroalkyl Acids. In *Toxicological Effects of Perfluoroalkyl and Polyfluoroalkyl Substances*, DeWitt, J. C., Ed. Springer International Publishing: Cham, 2015; pp 363-418.
2. Tsuda, S., Differential Toxicity between Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic Acid (PFOA). *J. Toxicol. Sci.* **2016,** *41*, SP27-SP36.
3. Barzen-Hanson, K. A.; Roberts, S. C.; Choyke, S.; Oetjen, K.; McAlees, A.; Riddell, N.; McCrindle, R.; Ferguson, P. L.; Higgins, C. P.; Field, J. A., Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater. *Environ. Sci. Technol.* **2017,** *51*, 2047-2057.
4. D'Agostino, L. A.; Mabury, S. A., Certain Perfluoroalkyl and Polyfluoroalkyl Substances Associated with Aqueous Film Forming Foam Are Widespread in Canadian Surface Waters. *Environ. Sci. Technol.* **2017,** *51*, 13603-13613.
5. Barzen-Hanson, K. A.; Field, J. A., Discovery and Implications of C_2 and C_3 Perfluoroalkyl Sulfonates in Aqueous Film-Forming Foams and Groundwater. *Environ. Sci. Technol Lett.* **2015,** *2*, 95-99.
6. Houde, M.; De Silva, A. O.; Muir, D. C. G.; Letcher, R. J., Monitoring of Perfluorinated Compounds in Aquatic Biota: An Updated Review. *Environ. Sci. Technol.* **2011,** *45*, 7962-7973.
7. Lang, J. R.; Allred, B. M.; Peaslee, G. F.; Field, J. A.; Barlaz, M. A., Release of Per- and Polyfluoroalkyl Substances (PFASs) from Carpet and Clothing in Model Anaerobic Landfill Reactors. *Environ. Sci. Technol.* **2016,** *50*, 5024-5032.
8. Sahu, S. P.; Qanbarzadeh, M.; Ateia, M.; Torkzadeh, H.; Maroli, A. S.; Cates, E. L., Rapid Degradation and Mineralization of Perfluorooctanoic Acid by a New Petitjeanite $Bi_3O(OH)(PO_4)_2$ Microparticle Ultraviolet Photocatalyst. *Environ. Sci. Technol. Lett.* **2018,** *5*, 533-538.
9. Merino, N.; Qu, Y.; Deeb, R. A..; Hawley, E. L.; Hoffmann, M. R.; Shaily, M., Degradation and Removal Methods for Perfluoroalkyl and Polyfluoroalkyl Substances in Water. *Environ. Eng. Sci.* **2016,** *33*, 615-649.
10. D'Agostino, L. A.; Mabury, S. A., Identification of Novel Fluorinated Surfactants in Aqueous Film Forming Foams and Commercial Surfactant Concentrates. *Environ. Sci. Technol.* **2014,** *48*, 121-129.
11. Xiao, X.; Ulrich, B. A.; Chen, B.; Higgins, C. P., Sorption of Poly- and Perfluoroalkyl Substances (PFASs) Relevant to Aqueous Film-Forming Foam (AFFF)-Impacted Groundwater by Biochars and Activated Carbon. *Environ. Sci. Technol.* **2017,** *51*, 6342-6351.
12. Bentel, M. J.; Yu, Y.; Xu, L.; Li, Z.; Wong, B. M.; Men, Y.; Liu, J., Defluorination of Per- and Polyfluoroalkyl Substances (PFASs) with Hydrated Electrons: Structural Dependence and Implications to PFAS Remediation and Management. *Environ. Sci. Technol.* **2019,** *53*, 3718-3728.
13. Hu, X. C.; Andrews, D. Q.; Lindstrom, A. B.; Bruton, T. A.; Schaider, L. A.; Grandjean, P.; Lohmann, R.; Carignan, C. C.; Blum, A.; Balan, S. A.; Higgins, C. P.; Sunderland, E. M., Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants. *Environ. Sci. Technol. Lett.* **2016,** *3*, 344-350.
14. Choi, Y. J.; Lazcano, R. K.; Yousefi, P.; Trim, H.; Lee, L. S., Perfluoroalkyl Acid Characterization in U.S. Municipal Organic Solid Waste Composts. *Environ. Sci. Technol. Lett.* **2019,** *6*, 372-377.
15. Singh, R. K.; Fernando, S.; Baygi, S. F.; Multari, N.; Thagard, S. M.; Holsen, T. M., Breakdown Products from Perfluorinated Alkyl Substances (PFAS) Degradation in a Plasma-Based Water Treatment Process. *Environ. Sci. Technol.* **2019,** *53*, 2731-2738.
16. Kononenko, I., Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. *Artif. Intell. in Med.* **2001,** *23*, 89-109.
17. LeCun, Y.; Bengio, Y.; Hinton, G., Deep Learning. *Nature* **2015,** *521*, 436-444.

18. Rosten, E.; Drummond, T., Machine Learning for High-Speed Corner Detection. In *Computer Vision - ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds. Springer: Berlin, Heidelberg, 2006; pp 430-443.
19. Pomerleau, D.; Jochem, T., Rapidly Adapting Machine Vision for Automated Vehicle Steering. *IEEE Expert* **1996,** *11*, 19-27.
20. Cherkasov, A.; Jonsson, M., A New Method for Estimation of Homolytic C−H Bond Dissociation Enthalpies. *J. Chem. Inf. Comput. Sci.* **2000,** *40*, 1222-1226.
21. Xue, C. X.; Zhang, R. S.; Liu, H. X.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T., An Accurate QSPR Study of O−H Bond Dissociation Energy in Substituted Phenols Based on Support Vector Machines. *J. Chem. Inf. Comput. Sci.* **2004,** *44*, 669-677.
22. Stanger, A., A Simple and Intuitive Description of C–H Bond Energies. *Eur. J. Org. Chem.* **2007,** *2007*, 5717-5725.
23. Feng, Y.; Liu, L.; Wang, J.-T.; Zhao, S.-W.; Guo, Q.-X., Homolytic C−H and N−H Bond Dissociation Energies of Strained Organic Compounds. *J. Org. Chem.* **2004,** *69*, 3129-3138.
24. Becke, A. D.; Johnson, E. R., A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005,** *123*, 154101.
25. Johnson, E. R.; Becke, A. D., A Post-Hartree–Fock Model of Intermolecular Interactions. *J. Chem. Phys.* **2005,** *123*, 024101.
26. Johnson, E. R.; Becke, A. D., A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006,** *124*, 174104.
27. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009,** *113*, 6378-6396.
28. Thomas, J.; Carrillo, M. J.; Serrato, A.; Lin, W.; Jäger, W.; Xu, Y., Rotational Spectroscopic and Theoretical Study of the Perfluorobutyric Acid···Formic Acid Complex. *J. Mol. Spectrosc.* **2017,** *335*, 88-92.
29. Thomas, J.; Seifert, N. A.; Jäger, W.; Xu, Y., A Direct Link from the Gas to the Condensed Phase: A Rotational Spectroscopic Study of 2,2,2-Trifluoroethanol Trimers. *Angew. Chem. Int. Ed. Engl.* **2017,** *56*, 6289-6293.
30. Nissen, J. H.; Stüker, T.; Drews, T.; Steinhauer, S.; Beckers, H.; Riedel, S., No Fear of Perfluorinated Peroxides: Syntheses and Solid-State Structures of Surprisingly Inert Perfluoroalkyl Peroxides. *Angew. Chem. Int. Ed. Engl.* **2019,** *58*, 3584-3588.
31. Baggioli, A.; Sansotera, M.; Navarrini, W., Thermodynamics of Aqueous Perfluorooctanoic Acid (PFOA) and 4,8-dioxa-3H-perfluorononanoic acid (DONA) from DFT Calculations: Insights Into Degradation Initiation. *Chemosphere* **2018,** *193*, 1063-1070.
32. Thomas, J.; Carrillo, M. J.; Serrato, A.; Xie, F.; Jäger, W.; Xu, Y.; Lin, W., Microwave Spectrum of the Complex of 3,3,3-trifluoro-2-(trifluoromethyl)propanoic Acid and Formic Acid. *Mol. Phys.* **2019,** *117*, 1193-1199.
33. DeWeerd, N. J.; Bukovsky, E. V.; Castro, K. P.; Kuvychko, I. V.; Popov, A. A.; Strauss, S. H.; Boltalina, O. V., Steric and Electronic Effects of CF_3 Conformations in Acene$(CF_3)_n$ Derivatives. *J. Fluorine Chem.* **2019,** *221*, 1-7.
34. Liu, J.; Van Hoomissen, D. J.; Liu, T.; Maizel, A.; Huo, X.; Fernández, S. R.; Ren, C.; Xiao, X.; Fang, Y.; Schaefer, C. E.; Higgins, C. P.; Vyas, S.; Strathmann, T. J., Reductive Defluorination of Branched Per- and Polyfluoroalkyl Substances with Cobalt Complex Catalysts. *Environ. Sci. Technol. Lett.* **2018,** *5*, 289-294.
35. Pari, S.; Wang, I. A.; Liu, H.; Wong, B. M., Sulfate Radical Oxidation of Aromatic Contaminants: A Detailed Assessment of Density Functional Theory and High-Level Quantum Chemical Methods. *Environ. Sci.: Processes Impacts* **2017,** *19*, 395-404.
36. Qu, X.; Latino, D. A. R. S.; Aires-de-Sousa, J., A Big Data Approach to the Ultra-Fast Prediction of DFT-Calculated Bond Energies. *J. Cheminf.* **2013,** *5*, 34.

37. Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. L., Recent Developments of the Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics. *Curr. Pharm. Des.* **2006,** *12*, 2111-2120.
38. Dahl, G. E.; Sainath, T. N.; Hinton, G. E. In *Improving Deep Neural Networks for LVCSR Using Rectified Linear Units and Dropout*, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 26-31, May 2013; pp 8609-8613.
39. Jones, L.; Whitaker, B. J., Modeling a Halogen Dance Reaction Mechanism: A Density Functional Theory Study. *J. Comput. Chem.* **2016,** *37*, 1697-1703.
40. Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L., The S66x8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated *Ab Initio* Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016,** *18*, 20905-20925.
41. Kesharwani, M. K.; Karton, A.; Martin, J. M. L., Benchmark *Ab Initio* Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016,** *12*, 444-454.
42. Hopmann, K. H., How Accurate is DFT for Iridium-Mediated Chemistry? *Organometallics* **2016,** *35*, 3795-3807.
43. Zarić, M. M.; Bugarski, B.; Kijevčanin, M. L., Best Methods for Calculating Interaction Energies in 2-butene and Butane Systems. *Comput. Theor. Chem.* **2017,** *1117*, 150-161.