

UC Davis

UC Davis Electronic Theses and Dissertations

Title

In Silico, In Vitro, and In Vivo Assessment of Trace Organic Contaminants (TrOCs)  
Bioaccumulation and Toxicity Potential

Permalink

<https://escholarship.org/uc/item/37v9094m>

Author

Li, Wenting

Publication Date

2022

Peer reviewed|Thesis/dissertation

In Silico, In Vitro, and In Vivo Assessment of Trace Organic Contaminants (TrOCs)  
Bioaccumulation and Toxicity Potential

By

WENTING LI  
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

---

Heather N. Bischel, Chair

---

Thomas M. Young

---

Sascha C.T. Nicklisch

Committee in Charge

2022

## ABSTRACT

An increasing numbers of trace organic contaminants (TrOCs) are being produced, used, and discharged into the environment. Over 17,000 pesticides and 4,730 per- and polyfluoroalkyl substances (PFAS) were reported by the Organisation for Economic Co-operation and Development (OECD) to be in circulation. Further TrOC diversity is generated when commercially produced compounds transform through myriad pathways. Understanding the public health impacts of environmental exposures to such a large number and wide structural diversity of TrOCs remains a significant challenge. Targeted chemical assays, the most common approach to bioaccumulation and risk assessment, provide a limited understanding of contaminant profiles in biological tissues and associated risks. More comprehensive analytical pipelines are needed. In this dissertation, **I develop and apply new methods to assess and predict bioaccumulation and toxicity of complex mixtures of TrOCs, using *in vivo*, *in vitro* and *in silico* approaches.** I present the three complementary strategies that together aim to improve management strategies for broad classes of TrOCs.

First, I develop and demonstrate an *in vivo* screening approach to assess the bioaccumulation and risk of TrOCs in complex mixtures. I apply the approach to assess pesticide and PFAS bioaccumulation in edible insect larvae (i.e., *H. illucens*) reared on agricultural by-products (i.e., almond hulls). Rather than targeting specific TrOCs from the onset, as is typical in chemical risk assessment, my approach broadly screens substrates and organism tissues for TrOCs using large databases of known contaminants. Multiple chemical extracts, obtained using methods that capture TrOCs with a broad range of physiochemical properties, are subjected to one or more of four analytical pipelines (two operational modes each for liquid and gas chromatography coupled with time-of-flight mass spectrometry). Semi-quantitative analysis of the substrates

screening revealed that bioaccumulative, persistent, and toxic chemicals are abundant in the agricultural by-products. Initial substrate screening was also used to guide targeted and non-targeted substrate and tissue analyses from *in vivo* bioaccumulation assays. Using this approach, I found that bifenthrin and a novel PFAS class bioaccumulate in the larvae tissue, initiating a pathway for contaminant transfer into the food chain.

Second, I use *in vitro* and *in silico* approaches to assess bioaccumulation potential of PFAS in a commercially available class-B firefighting foam—aqueous film forming foams (AFFF). PFAS accumulation in serum and organ tissues are caused by PFAS precursor metabolisms and protein bindings. It is challenging to identify and quantitatively assess the contribution of each bioaccumulation pathway. However, mechanistic understanding in PFAS bioaccumulation is essential for prediction model and bioaccumulation potential assessment across the chemicals. In this dissertation, I collect binding data from equilibrium dialysis experiments using human serum albumin (HSA), the most abundant serum protein in human blood, and AFFF. This *in vitro* approach eliminates the influence of biotransformation to bioaccumulation and help discover ultra-strong binding or potentially covalent binding PFAS to HSA. I further use experimental data to test the effectiveness of molecular docking predicted HSA-binding and -nonbinding compounds in AFFF. The combination of *in vitro* and *in silico* approaches provide replicable, high-throughput workflows for assessing bioaccumulation potentials of chemicals in commercial products that are structurally diverse like PFAS.

Third, I use *in silico* simulations validated in Part 2 to generate noncovalent binding scores of over 4,760 PFAS structures to human carrier and receptor proteins. In this dissertation, I hypothesize that protein binding scores generated from molecular docking can improve model performance for bioactivity/toxicity prediction of PFAS. Moreover, the prediction model can be

used to predict bioactivity/toxicity of PFAS with similar chemical structure. I test this theory by correlating protein binding scores along with 45 other chemical descriptors to negative health endpoints using linear discrimination analysis and machine learning algorithms. This quantitative structure-activity relationship (QSAR) uses state-of-the-art knowledge to predict toxicity of PFAS with known structure. Altogether, the combination of experimental and modeling techniques provided value in assessing the bioaccumulation and toxicity of organic contaminants.

## **DEDICATION**

*This dissertation is dedicated to my parents and 'Still us 3' group chat members for all the love, support, and encouragement through this wonderful journey.*

## ACKNOWLEDGEMENTS

My doctoral journey is filled with difficult challenges, rich learning experiences, and hilarious moments. After facing countless failures and mistakes, I was able to find the nuggets of wisdom and navigate my way through, because of the love, support, encouragement, guidance, and help I received. I have been telling everyone that I had the best time of my life in the graduate program. It has transformed me to a person I would have never imagined. Here, I would like to thank all the people who have been a part of my doctoral journey.

I would love to thank Prof. Heather Bischel. I never thought I could get a PhD before I met Heather. When I first started my master, I was so confused and uncertain about what I could and what I liked to do. I was calling my mom one month into the master program thinking about quitting because without an engineering background I was totally overwhelmed by the amount of course work in engineering school. Heather's enthusiasm and encouragement sparked the fire in me for science and gave me the confidence, resource, and freedom to explore and truly enjoying the process of researching. I feel so lucky to have you as my PI, Heather. You helped me find what I loved doing, which might take one's whole life to search for.

Meanwhile, Bischel group has provided me tremendous support in research and life in general. I love our weekly group meetings. Although most of us work on completely different topics, interesting discussions helped me think out of the box every time. Olivia Wrightwood, thank you, for being my best friend, for walking with me through all the happy and sad time for the past five years, for being an inspirational engineer, and for being the best aunt of Peipei. Without your and Laura Wrightwood's love and support, I couldn't have done it. Thank you, Dr. Camille Wolf, for being such an amazing colleague and friend. You are my partner in crime, in lab, in business, and in the gym. I cannot wait to conquer another mountain with you! Thank you, Madison Hattaway, for helping me with nontarget analysis, searching for my lost wallet in Tempe, AZ, guiding me through my first marathon, and letting me steal your snacks through all these hot days in the office. Thank you, Hannah Safford, for generously reviewing my manuscripts and proposals and for being such an inspirational and empowering figure in the group. Thank you, Mel Johnson, for being a part of my endurance journey (never forget Chico Go Big or Go Home, 2022) and for valuable discussions of water treatment in disadvantaged communities and entrepreneurship. Thank you Michelle Marie Clauzel, Bischel group's master student, lab

manager, and dearest friend, for your help with all the bureaucracy for instrument and material orders. Thank you to our undergraduate researchers, Yuhong Hu, Abby Shibiru, Eduardo Baptista, for helping me with various tasks.

I would love to thank Prof. Thomas Young's research group. Thank you Tom, for your countless guidance, advice, edits, and instrumentation resource! This dissertation is only possible because of your group's support. Dr. Gabrielle Black, Luann Wong, and Chris Alaimo, thank you for your help with method development, LC and GC acquisitions, and software support.

I would also love to thank the black soldier fly larvae team at Biological Agricultural Engineering department. Prof. Jean VanderGheynst, thank you for including me as a part of this project and providing insights in experimental design. Dr. Lydia Palma and Dr. Jesus Dionisio Fernandez Bayo, thank you for the help of setting up bioreactors and sorting larvae pre- and post-incubation. Dr. Trevor M Fowles, thank you for providing all the initial larvae used in this study.

Additionally, I have received a lot of help from the academia outside of UC Davis. Dr. Carrie McDonough (Stony Brook University) and Prof. Jamie Dewitt (East Carolina University) thank you for including me as a part the perspective paper and providing your expert opinions on my PFAS protein binding and QSAR projects. Prof. Christopher Higgins (Colorado School of Mines), thank you for the AFFF sample. Prof. Carla Ng (University of Pittsburgh), thank you for providing your expert opinions on molecular simulations and QSAR models.

Most indispensably, I would like to thank my families. Mom (Yanying), and dad (Peigang), it has been a difficult five-year for us as a family (covid certainly did not help). Although we only get to see each other twice for the past five years, I know that you try everything you can to help and support me. I want you guys to know that you have done more than enough. I can't ask for more. I wish I was home for you guys when Laolao passed away and Laoye got sick. Thank you for tolerating my stinky attitude and spontaneous adventures from time to time. Hopefully, we will reunite soon!

## LIST OF PAPERS

This dissertation is based on three individual scientific research and one perspective papers.

### Paper I

**Li, W.;** Hu, Y.; Bischel, H.N. In-Vitro and In-Silico Assessment of Per- and Polyfluoroalkyl Substances (PFAS) in Aqueous Film-Forming Foam (AFFF) Binding to Human Serum Albumin. *Toxics* 2021, 9, 63. <https://doi.org/10.3390/toxics9030063>

### Paper II

McDonough, C. A.; **Li, W.;** Bischel, H. N.; De Silva, A. O.; and DeWitt, J. C. Widening the Lens on PFASs: Direct Human Exposure to Perfluoroalkyl Acid Precursors (pre-PFAAs). *Environmental Science & Technology* Article, <https://doi.org/10.1021/acs.est.2c00254>

### Paper III

**Li, W. &** Bischel, H. N. (2022). Are resource recovery insects safe for feed and food? A screening approach for bioaccumulative trace organic contaminants. *Science of The Total Environment*, 837, 155850, <https://doi.org/10.1016/j.scitotenv.2022.155850>

### Paper IV

**Li, W.;** Bischel, H.N. Development of a PFAS toxicity prediction model using QSAR tools and bioactivity data. (Manuscript in preparation).

## **STATEMENT OF RESPONSIBILITIES**

My contributions to the four papers included in this thesis were:

### **Paper I and III**

I have performed all analytical lab work regarding method development, optimization, and validation, sample collection, extraction, and analysis. I was also responsible for data acquisition, interpretation and took the lead in writing the paper.

### **Paper III**

I contributed to this perspective paper in bioaccumulation and analytical part of the original manuscript writing and overall editing. None of the text from this published manuscript was directly used in this dissertation. However, the same knowledge gap and overlooked issues of PFAS are mentioned in the introduction of this dissertation.

### **Paper IV**

I contributed to the planning of this study. I curated the bioactivity database, performed simulation of all the models used, and performed statistical tests to evaluate the testing models. I also interpreted the data and took the lead in writing the paper.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>ii</b>
<b>DEDICATION</b> .....	<b>v</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>vi</b>
<b>LIST OF PAPERS</b> .....	<b>viii</b>
<b>STATEMENT OF RESPONSIBILITIES</b> .....	<b>ix</b>
<b>TABLE OF CONTENTS</b> .....	<b>x</b>
<b>LIST OF TABLES</b> .....	<b>xiii</b>
<b>LIST OF FIGURES</b> .....	<b>xiii</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
<b>1.1 Organic chemicals of emerging concerns</b> .....	<b>1</b>
<b>1.2 Analytical techniques used for TrOCs</b> .....	<b>2</b>
<b>1.3 PFAS as a case study</b> .....	<b>3</b>
<b>1.4 Knowledge gap addressed in each chapter</b> .....	<b>3</b>
<b>1.5 Reference</b> .....	<b>5</b>
<b>CHAPTER 2: ARE RESOURCE RECOVERY INSECTS SAFE FOR FEED AND FOOD? A SCREENING APPROACH FOR BIOACCUMULATIVE TRACE ORGANIC CONTAMINANTS.</b> .....	<b>7</b>
<b>2.1 Abstract</b> .....	<b>8</b>
<b>2.2 Introduction</b> .....	<b>9</b>
<b>2.3 Method and material</b> .....	<b>11</b>
2.3.1 <i>Experimental design</i> .....	11
2.3.2 <i>Feedstock source and characterization</i> .....	12
2.3.3 <i>Larvae bioaccumulation reactors</i> .....	14
2.3.4 <i>Larvae tissue extraction</i> .....	14
2.3.5 <i>HPLC-MS analysis</i> .....	15
2.3.6 <i>GC-MS analysis</i> .....	15
2.3.7 <i>Spike recovery test</i> .....	16
2.3.8 <i>Data processing and analysis</i> .....	16
<b>2.4 Results and discussion</b> .....	<b>17</b>
2.4.1 <i>Qualification of residual pesticides and PFAS in almond hulls</i> .....	17
2.4.2 <i>Quantification of pyrethroids in almond hulls</i> .....	21
2.4.3 <i>Bioaccumulation of bifenthrin in BSFL</i> .....	22
2.4.4 <i>Uptake of spiked PFAS targets in BSFL</i> .....	24

2.4.5 Nontarget analysis of PFAS in almond hulls .....	25
2.4.6 Nontarget analysis reveals uptake of novel PFAS in BSFL .....	26
<b>2.5 Conclusion.....</b>	<b>29</b>
<b>2.6 References .....</b>	<b>30</b>
<b>CHAPTER 3: IN-VITRO AND IN-SILICO ASSESSMENT OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFAS) IN AQUEOUS FILM-FORMING FOAM (AFFF) BINDING TO HUMAN SERUM ALBUMIN .....</b>	
<b>3.1 Abstract.....</b>	<b>37</b>
<b>3.2 Introduction .....</b>	<b>38</b>
<b>3.3 Method and material.....</b>	<b>40</b>
3.3.1 Study Design and Workflow .....	40
3.3.2 Equilibrium Dialysis .....	41
3.3.3 PFAS Extractions.....	42
3.3.4 Analytical Instrumental Set-up .....	42
3.3.5 Suspect Screening .....	43
3.3.6 Targeted MS/MS.....	44
3.3.7 PFAS Quantification .....	44
3.3.8 Experimental Determination of PFAS Noncovalent Binding Affinities .....	44
3.3.9 Computational Simulations of Noncovalent PFAS Protein Binding.....	47
<b>3.4 Results and discussion.....</b>	<b>48</b>
3.4.1 Characterization of PFAS in AFFF .....	48
3.4.2 Noncovalent Binding of PFAS in AFFF to Human Serum Albumin.....	49
3.4.3 Residual PFAS in Precipitated Protein Pellet Covalently Binding to HSA .....	52
3.4.4 Quantitative Determination of PFAS–HSA Association Constants.....	55
3.4.5 Evaluation of Molecular Docking to Predict the PFAS–HSA Binding Affinities .....	55
3.4.6 Semi-quantification of Bioconcentration Factors of Qualified PFAS.....	57
3.4.7 Qualitative Prediction of HSA-bound vs. Nonbound Compounds .....	60
<b>3.5 Conclusion.....</b>	<b>63</b>
<b>3.6 References .....</b>	<b>64</b>
<b>CHAPTER 4: DEVELOPMENT AND PREDICTION OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFAS) TOXICITY USING MOLECULAR DOCKING AND QSAR TOOLS .....</b>	
<b>4.1 Introduction.....</b>	<b>70</b>
<b>4.2 Method .....</b>	<b>73</b>
4.2.1 Computational resource and machine learning packages .....	73
4.2.2 Chemical database curation.....	73
4.2.3 Bioactivity database curation .....	74
4.2.4 QSAR model tools.....	76
4.2.5 Model optimization and validation.....	76
<b>4.3 Result and discussion.....</b>	<b>78</b>
4.3.1 Identification of significant chemical descriptors .....	78
4.3.2 Performance of QSAR models .....	82

<i>4.3.3 Correlating PFAS structural analysis with bioactivity/toxicity classification</i> .....	84
<b>4.4 Reference</b> .....	<b>87</b>
<b>CHAPTER 5: CONCLUDING REMARK</b> .....	<b>92</b>
<b>APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2</b> .....	<b>95</b>
<b>APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3</b> .....	<b>116</b>
<b>APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4</b> .....	<b>194</b>

## LIST OF TABLES

- Table 1** List of descriptors used in development of QSAR model
- Table 2** Summary of the five toxicity/bioactivity datasets used to build QSAR model

## LIST OF FIGURES

- Figure 1** Schematic illustrating three stages of our experimental design to investigate bioaccumulation of pesticides and PFAS in BSFL.
- Figure 2** Persistent, bioaccumulative, and relatively more toxic compounds identified in six almond hulls.
- Figure 3** Concentrations of pyrethroid pesticides quantified in Hulls #3 and #5. Almond hulls were used as substrates for investigating pyrethroid bioaccumulation in black soldier fly larvae (BSFL).
- Figure 4** Bifenthrin concentrations measured in almond hulls and black soldier fly larvae (BSFL).
- Figure 5** Kendrick mass defect (KMD) analysis plots of Hull #5 that was spiked with 24 PFAS standards and their relative abundance in hulls and larvae.
- Figure 6** Equilibrium dialysis set-up and observed mass balance for three PFAS.
- Figure 7** Experimentally determined noncovalent association constants ( $K_A$ ) of 26 PFAS targets with human serum albumin (HSA) in equilibrium dialysis.
- Figure 8** Residual PFAS quantified in digested HSA pellets, precipitated from equilibrium dialysis experiment.

- Figure 9** Comparison of experimental  $\log K_A$  with results from molecular simulations with the HSA crystal structure 1E7G.
- Figure 10** Pseudo-bioconcentration factors ( $BCF_{i,pseudo}$ ) of PFAS in aqueous film-forming foams (AFFF) that were bound noncovalently (yellow) or potentially covalently (red) to HSA.
- Figure 11** Violin plots of molecular docking simulated 1E7G docking scores for PFAS and hydrocarbon surfactants identified in AFFF.
- Figure 12** Receiver Operating Characteristic (ROC) curve of SFS-LDA
- Figure 13** Predicting performance of training and validation datasets in Area Under The Curve - Receiver Operating Characteristics (AUC-ROC) scores and Matthews Correlation Coefficient (MCC) of different testing models.
- Figure 14** Structure-Activity Landscape Index (SALI) activity cliff analysis (ACA) for input PFAS and manufactured and used OECD PFAS in PubChem.

## CHAPTER 1: INTRODUCTION

### 1.1 Organic chemicals of emerging concerns

The number of chemicals manufactured and commercially available has grown exponentially over the past two decades, increasing from 20 million in 2002 to 198 million in 2022 according to the Chemical Abstract Service Registry (CAS) (<https://www.cas.org/cas-data/cas-registry>; accessed on August 9<sup>th</sup> 2022). Many such chemicals are essential to modern life, but some anthropogenic organic chemicals are produced with little safety assessment. Disposal of used products and incomplete treatment of problematic chemicals into water and wastewater systems results in a complex mélange of chemicals in the environment. Trace organic contaminants (TrOCs) are ubiquitous in water, soil, air, and biota (Logemann et al., 2022). This can lead to undesirable and sometimes dangerous effects on ecosystems and public health (Diamond et al., 2015; Escher et al., 2020). Moreover, chemical regulations are often reactionary in nature, addressing chemical threats long after their discovery. As legacy contaminants phase out of the market, new chemicals (e.g., replacements) are developed, further complicating analysis of complex mixtures of TrOCs.

Current chemical management strategies fail to address safety concerns at the global scale. The cost of generating safer replacement chemicals and clean-up is often too high (e.g., long time horizons and large development costs) relative to business-as-usual. As a result, production and management of hazardous chemicals are “locked in,” creating a dependency mechanism akin to our continued use of fossil fuels despite the known harmful effects of climate change. Despite strong negative impacts of a sub-class of chemicals, the supply of hazardous chemicals continues driven by the demand (Blumenthal et al., 2022; Melymuk et al., 2022). So how do we prevent the system locking in? How do we open the lock when the system is already established? In this

dissertation, I seek to (1) develop more efficient chemical risk assessment workflows to provide information before hazardous chemical production infrastructure is established and (2) guide novel chemical production based on biological activity of existing chemicals.

## **1.2 Analytical techniques used for TrOCs**

High-resolution mass spectrometry (HRMS) is a powerful tool that combines analytical sensitivity and mass accuracy (<5 ppm) allowing simultaneously study known and newly identified TrOCs. There are three major analytical approaches for HRMS: target, suspect-screening, and nontarget analyses. Among three approaches, targeted analysis is the most sensitive method, which identifies and quantifies a small number (e.g., 10s) of targeted TrOCs using known mass-to-charge ratio of the ions and their retention time (Pourchet et al., 2020). Suspect-screening and nontarget approaches in comparison, acquire full scans of molecular ions and their fragmentation ions, which allow the analysis of greater number of chemicals. However, background ions that coelute with ions of interest generally lower instrumental sensitivity. In suspect screening, chemical information (e.g., mass spectra) acquired from an environmental sample is screened against an extensive chemical spectra library. Based on accurate mass, fragmentation, coelution, and isotopic patterns, suspect-screening can identify and semi-quantitatively assess an extensive list of substances (e.g., 1000s) when analytical standard is not available (Y. Wang et al., 2018). Nontarget analysis techniques have been popularized in recent year for complex environmental samples with little to no chemical information. In nontarget analysis, large number of chemical features (e.g., 100,000s) can often be extracted from one sample (Moschet et al., 2018a). To prioritize compounds that may require identification (e.g., the novel ones that pose environmental and/or public health concerns), it is critical and challenging to perform statistical analyses such as principal component analysis, clustering, and regression

analysis for data reduction (Knolhoff & Fisher, 2021). In this dissertation, I use HRMS with all three analytical approaches to study known and newly identified TrOCs.

### **1.3 PFAS as a case study**

Poly- and perfluoroalkyl substances (PFAS) are good examples of locked-in TrOCs. PFAS were firstly developed during 1940s as a part of the Manhattan project (Harsanyi & Sandford, 2015), prior to the establishment of United States Environmental Protection Agency (EPA) and the Toxic Substances Control Act of 1976. PFAS have many favorable properties for commercial applications. PFAS are highly persistent with high thermal and chemical stability due to the presence of abundant C-F bonds, which are the strongest single bonds in organic chemistry. Additionally, PFAS are better surfactants compared to their hydrocarbon counterparts due to the strong electronegativity and small atomic size of fluorine, which enhance surface activity at very low concentrations (e.g., amphiphilicity or amphiphobicity) (Z. Wang et al., 2017). Based on these favorable properties, PFAS are widely used in commercial products, such as non-stick cookware, food packaging, personal products, pesticide formulations, textiles, aqueous film-forming foams (AFFF) (Glüge et al., 2020). However, these unique chemical properties also make PFAS highly persistent and mobile. Some PFAS are also known to be bioaccumulative and toxic. Exposure to many PFAS has been linked to diseases such as cancer, cardiovascular disease, kidney disease, liver disease, infectious disease, neurological disease, type II diabetes, osteoarthritis, immune suppression, and respiratory disease (ATSDR, 2017).

### **1.4 Knowledge gap addressed in each chapter**

Bioaccumulation of chemicals in an organism that may be toxic is of major environmental concern. Though organisms in the environment are often exposed to mixtures of TrOCs, most bioaccumulation studies use targeted approaches. With little information for the contaminants and

the chemical nature of sample matrices, assessing bioaccumulation potentials of mixtures of TrOCs in organism tissue is challenging. In chapter 2, I address this knowledge gap by building a screening based chemical analysis workflow. The workflow screens TrOCs (e.g., pesticides, pharmaceuticals, plasticizer, etc.) in nutrient recovery insects reared from agricultural waste. TrOCs are then prioritized based on “persistent, bioaccumulative, and toxic (PBT)” (ECHA, 2017) and their relative abundance, which also narrow down the type of analytical technique for the downstream bioaccumulation study. In this chapter, I also explore the use of nontarget analysis for identifying novel chemicals that may accumulate in the model insect.

Studies indicate that interactions of PFAS with carrier proteins are important determinants of PFAS bioaccumulation (Ng & Hungerbühler, 2014). Unfortunately, these relationships have been established experimentally for only a small subset of PFAS with little relevant matrix effect (e.g., studies use protein with concentrations magnitudes lower or PFAS with concentrations magnitudes higher than biological relevant levels), due to the costly and inefficient nature of targeted analyses. More comprehensive analytical pipelines are needed. Therefore, I use a suspect-screening approach to assess PFAS binding affinities and mechanisms with human serum albumin (HSA) protein at biologically and environmentally relevant exposure levels in Chapter 3. I further explore the use of molecular docking to predict HSA binding affinities.

Increasing numbers of PFAS have been produced, and advancements in analytical techniques are revealing extreme diversity of PFAS structures. Additionally, OECD broadened the definition of what constitutes a PFAS. Considering these factors, the number of PFAS detected in the environment and/or applied in commercial products has increased from 3,000 to 6 million (OECD PFAS listed in PubChem) over the last 5 years (i.e. my doctoral journey). I recognize the large gap between data and urgent regulatory needs from the public. We need to extrapolate our

knowledge on known PFAS to less known PFAS using data science (e.g., data mining and machine learning). In Chapter 4, I develop a PFAS toxicity quantitative structure activity relationship (QSAR) model and investigate the importance of docking scores of PFAS to HSA, fatty acid binding proteins (FABPs), and peroxisome proliferator-activated receptors (PPARs) as descriptors.

## 1.5 Reference

- ATSDR. (2017). *An Overview of Perfluoroalkyl and Polyfluoroalkyl Substances and Interim Guidance for Clinicians Responding to Patient Exposure Concerns Interim Guidance*. [https://www.atsdr.cdc.gov/pfas/docs/pfas\\_clinician\\_fact\\_sheet\\_508.pdf](https://www.atsdr.cdc.gov/pfas/docs/pfas_clinician_fact_sheet_508.pdf)
- Blumenthal, J., Diamond, M. L., Hoffmann, M., & Wang, Z. (2022). Time to Break the “lock-In” Impediments to Chemicals Management. In *Environmental Science and Technology* (Vol. 56, Issue 7, pp. 3863–3870). American Chemical Society. <https://doi.org/10.1021/acs.est.1c06615>
- Diamond, M. L., de Wit, C. A., Molander, S., Scheringer, M., Backhaus, T., Lohmann, R., Arvidsson, R., Bergman, Å., Hauschild, M., Holoubek, I., Persson, L., Suzuki, N., Vighi, M., & Zetzsch, C. (2015). Exploring the planetary boundary for chemical pollution. In *Environment International* (Vol. 78, pp. 8–15). Elsevier Ltd. <https://doi.org/10.1016/j.envint.2015.02.001>
- ECHA. (2017). Guidance on Information Requirements and Chemical Safety Assessment (Chapter R.11: PBT/vPvB assessment, Version 3.0). In *EUROPEAN CHEMICALS AGENCY (ECHA)* (Issue June). <https://doi.org/10.2823/128621>
- Escher, B. I., Stapleton, H. M., & Schymanski, E. L. (2020). Tracking complex mixtures of chemicals in our changing environment. *Science*, 367(6476), 388–392. [www.norman-network.com/nds/](http://www.norman-network.com/nds/)
- Glüge, J., Scheringer, M., Cousins, I. T., Dewitt, J. C., Goldenman, G., Herzke, D., Lohmann, R., Ng, C. A., Trier, X., & Wang, Z. (2020). An overview of the uses of per- And polyfluoroalkyl substances (PFAS). *Environmental Science: Processes and Impacts*, 22(12), 2345–2373. <https://doi.org/10.1039/d0em00291g>
- Harsanyi, A., & Sandford, G. (2015). Organofluorine chemistry: Applications, sources and sustainability. *Green Chemistry*, 17(4), 2081–2086. <https://doi.org/10.1039/c4gc02166e>
- Knolhoff, A. M., & Fisher, C. M. (2021). Strategies for data reduction in non-targeted screening analysis: The impact of sample variability for food safety applications. *Food Chemistry*, 350. <https://doi.org/10.1016/j.foodchem.2020.128540>
- Logemann, A., Reininghaus, M., Schmidt, M., Ebeling, A., Zimmermann, T., Wolschke, H., Friedrich, J., Brockmeyer, B., Pröfrock, D., & Witt, G. (2022). Assessing the chemical anthropocene – Development of the legacy pollution fingerprint in the North Sea during the last century. *Environmental Pollution*, 302. <https://doi.org/10.1016/j.envpol.2022.119040>
- Melymuk, L., Blumenthal, J., Sáňka, O., Shu-Yin, A., Singla, V., Šebková, K., Pullen Fedinick, K., & Diamond, M. L. (2022). Persistent Problem: Global Challenges to Managing PCBs. *Environmental Science & Technology*, 56(12), 9029–9040. <https://doi.org/10.1021/acs.est.2c01204>
- Moschet, C., Anumol, T., Lew, B. M., Bennett, D. H., & Young, T. M. (2018). Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution

- Mass Spectrometric Study. *Environmental Science and Technology*, 52(5), 2878–2887. <https://doi.org/10.1021/acs.est.7b05767>
- Ng, C. A., & Hungerbühler, K. (2014). Bioaccumulation of perfluorinated alkyl acids: Observations and models. *Environmental Science and Technology*, 48(9), 4637–4648. <https://doi.org/10.1021/es404008g>
- Pourchet, M., Debrauwer, L., Klanova, J., Price, E. J., Covaci, A., Caballero-Casero, N., Oberacher, H., Lamoree, M., Damont, A., Fenaille, F., Vlaanderen, J., Meijer, J., Krauss, M., Sarigiannis, D., Barouki, R., le Bizec, B., & Antignac, J. P. (2020). Suspect and non-targeted screening of chemicals of emerging concern for human biomonitoring, environmental health studies and support to risk assessment: From promises to challenges and harmonisation issues. *Environment International*, 139. <https://doi.org/10.1016/j.envint.2020.105545>
- Wang, Y., Yu, N., Zhu, X., Guo, H., Jiang, J., Wang, X., Shi, W., Wu, J., Yu, H., & Wei, S. (2018). Suspect and Nontarget Screening of Per- and Polyfluoroalkyl Substances in Wastewater from a Fluorochemical Manufacturing Park. *Environmental Science and Technology*, 52(19), 11007–11016. <https://doi.org/10.1021/acs.est.8b03030>
- Wang, Z., Dewitt, J. C., Higgins, C. P., & Cousins, I. T. (2017). A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environmental Science and Technology*, 51(5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>

**CHAPTER 2: ARE RESOURCE RECOVERY INSECTS SAFE FOR FEED AND FOOD?  
A SCREENING APPROACH FOR BIOACCUMULATIVE TRACE ORGANIC  
CONTAMINANTS.**

Wenting Li<sup>1</sup>, Heather N. Bischel\*<sup>1</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, University of California Davis, California,  
95616, United States

Published – Science of the Total Environment

Radarweg 29 Amsterdam 1043 NX, Netherlands

<https://doi.org/10.1016/j.scitotenv.2022.155850>

## 2.1 Abstract

Most bioaccumulation assessments select one or several compound classes *a priori* for analysis performed by either liquid or gas chromatography coupled with mass spectrometry (LC-MS or GC-MS). When organisms are exposed to complex mixtures of trace organic contaminants (TrOCs), targeted chemical assays limit understanding of contaminant profiles in biological tissues and associated risks. We used a semi-quantitative suspect-screening approach to assess the bioaccumulation potential of diverse TrOCs in black soldier fly larvae (BSFL) using almond hulls (by-products of the almond industry in California) as test substrates. BSFL digestion is gaining traction as a resource recovery strategy to generate animal feed from low-value organic wastes. We screened almond hulls from six California farms for the presence of 5,728 TrOCs using high resolution mass spectrometry. We then categorized the risk potential of 46 TrOCs detected in the hulls based on their predicted bioaccumulation, persistence, and toxicity in order to select two hulls for an *in situ* BSFL bioaccumulation screening study. We analyzed larvae tissues and feeding substrate initially and after 14 days of growth using targeted, suspect-screening, and nontarget-screening methods. The survival rate of BSFL in all rearing reactors was greater than 90%, indicating low toxicity of the substrates to BSFL. Esfenvalerate, cyhalothrin, and bifenthrin were the most abundant insecticides quantified (81.7 to 381.6 ng/g-dw) in the hulls. Bifenthrin bioaccumulated in BSFL tissues (14-day bioaccumulation factor, BAF, of  $2.17 \pm 0.24$ ). For nontarget analysis, Kendrick mass defect (KMD) analysis of PFAS homologous series revealed hydrogen-substituted perfluoroalkyl carboxylic acids (H-PFCAs) in the hulls and BSFL tissues after growth. Our approach demonstrates the utility of suspect-screening in chemical safety assessments when organic wastes with highly diverse and variable contaminant profiles are used in resource recovery pipelines.

## 2.2 Introduction

Black soldier fly larvae (BSFL) are proposed as ecologically favorable replacements for traditional animal feeds (e.g., fishmeal and soymeal) to help address projected global food shortages (Fowles & Nansen, 2019; Prosekov & Ivanova, 2018; Shumo et al., 2019). Large numbers of BSFL can be farmed using organic wastes as substrates (Kumar et al., 2018; Perednia et al., 2017; Salomone et al., 2017; Win et al., 2018). BSFL can then be processed into marketable end products such as biofuels (using fatty insect tissues) or higher-value soil amendments (enriched with chitin from insect castings) (Zheng et al., 2012; Surendra et al., 2016; Gold et al., 2018). BSFL are easy to handle and have a relatively short life-cycle (~ 4 weeks), facilitating large-scale production (Fowles & Nansen, 2019). Numerous companies are already capitalizing on the capacity of insects to digest food and agricultural waste to generate value (Filou, 2021; Fowles & Nansen, 2019; Gold et al., 2018).

Production of BSFL on almond hulls as substrates could be beneficial for both organics waste management and nutrient recovery (Palma et al., 2018, 2019). California supplies approximately 80% of the world's almonds, generating 2.55 million tons of hulls annually (CALIFORNIA ALMOND INDUSTRY FACTS, 2016). Hulls are often used in feedstock for dairy cows as substitutes for mid-grade alfalfa (Williams et al., 2018; Hart et al., 2020). As production costs have gradually moved the dairy industry to other states (Hatzenbuehler et al., 2021), alternative uses of hulls are needed in California. Several studies have demonstrated the economic potential of using BSFL for almond hull waste reduction and nutrient recovery (Palma et al., 2019; Shumo et al., 2019; Gold, Marie, et al., 2020; Soldier et al., 2021). BSFL have much wider applications as feed (e.g., feed for chicken, fish, swine, etc) compared to almond hulls, which have high fiber content but limited digestibility (Miller et al., n.d.). Alongside potential economic

gains, consideration of the impacts of agricultural pesticides on production efficiency and end-product quality are important when insects such as BSFL are used to generate animal feed from agricultural organic wastes. Almond orchards usually apply insecticides at (1) the beginning of hull split and (2) when hulls have completely opened (X. Li et al., 2021). Use of pyrethroid insecticides on almonds in California was associated with environmental contamination in the San Joaquin River watershed (X. Zhang et al., 2008; Zhan & Zhang, 2014). Pyrethroids were popularized in 1980s to replace organophosphates due to their relatively low acute mammalian toxicity (Costa, 2015), but they are now known to cause detrimental effects to non-target species including fish and other aquatic species (Bhatt et al., 2019).

Production of BSFL on pre- or post-consumer food waste including manure, kitchen waste, and agricultural byproducts further triggers concerns for food safety (Nguyen et al., 2015). Trace organic contaminants (TrOCs) such as pyrethroids may bioaccumulate in larval tissue during digestion of the substrates, initiating a pathway for contaminant transfer into the food chain. In addition to pesticides, other TrOCs in agricultural byproducts are likely to be present and of potential concern. For instance, poly- and perfluoroalkyl substances (PFAS) were found in commercial pesticide products either as inert ingredients or contaminants due to fluorinated polymers use in storage containers (Glüge et al., 2020; Kauffman, 2021; US Environmental Protection Agency, 2021). Realistic assessment of the chemical safety of products derived from organic wastes is complicated by the diversity and variability of contaminants in waste mixtures. For PFAS alone, there were over 8,000 PFAS structures identified in over 200 applications (Elich et al., 2022), most of which do not have analytical standards for identification or quantification. Bioaccumulation studies that target a limited number of chemicals (e.g., <10 mycotoxins or insecticides (Purschke et al., 2017; Id et al., 2021)) will do little to assess the complete risk profiles

of BSFL grown from waste mixtures. Current analytical and regulatory strategies for detecting chemicals one-by-one are also costly, incomplete, and time consuming (Hollender et al., 2019; Lara-Martín et al., 2020).

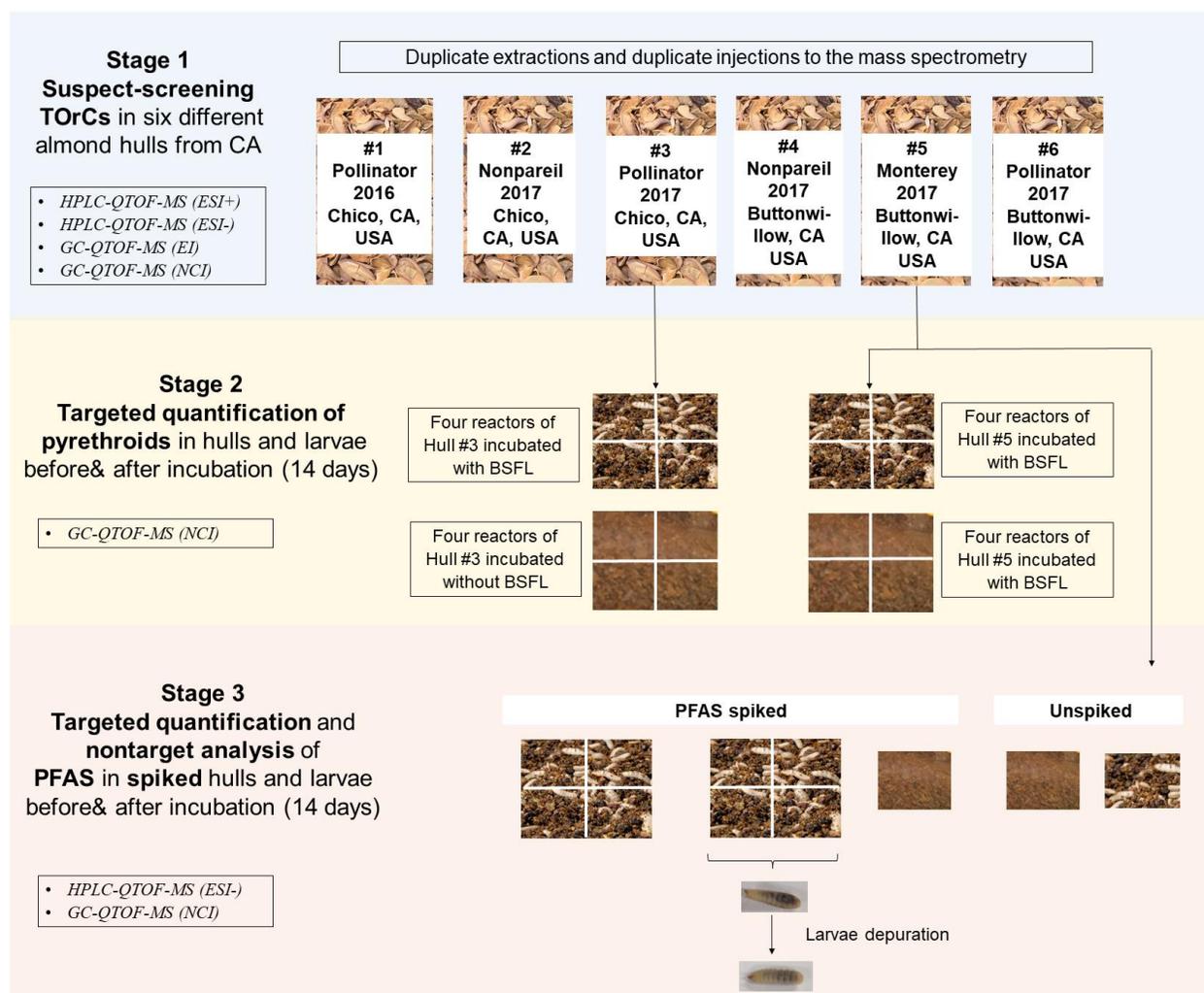
In this study we used the BSFL-hull model system and high-resolution mass spectrometry (HRMS) to (1) characterize environmental TrOCs in almond hulls, (2) develop a workflow for semi-quantitative chemical screening and prioritization of environmental TrOCs in organic wastes, and (3) quantify pyrethroids and PFAS in BSFL produced with almond hulls as substrate.

## **2.3 Method and material**

### *2.3.1 Experimental design*

This study progressed in three stages (Figure 1). In Stage 1, we analyzed organic solvent extracts of six types of almond hulls (hereafter referred to as “hulls”) using a suite of quadrupole time of flight high resolution mass spectrometry (QTOF-HRMS) pipelines. Extracts were analyzed using both gas chromatography (GC) and high-performance liquid chromatography (HPLC). We ran HPLC with positive (ESI+) and negative (ESI-) electrospray ionization modes, and GC with electron ionization (EI) and negative chemical ionization (NCI) modes. Data acquired from multiple collision energies was screened against water contaminant mass spectra libraries that included pesticides and PFAS, amongst other TrOCs. In Stage 2, we conducted a bioaccumulation screening study using BSFL. We selected two of the original six hull types that contained the most abundant, bioaccumulative, persistent, and toxic pesticides. We quantified pyrethroids that accumulated in BSFL and that remained in the residual feedstock following BSFL digestion. In Stage 3, we grew BSFL on one hull type that was unspiked or spiked with five PFAS—perfluoro-n-butanoic acid (PFBA), perfluoro-n-octanoic acid (PFOA), perfluoro-1-butanefulfonic acid

(PFBS), linear perfluorooctanesulfonic acid (L-PFOS), and ammonium 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propanoate (GenX)—to obtain quantifiable level of PFAS in BSFL and residual hulls. We analyzed the hulls, BSFL tissues, and residual hulls using GC- and HPLC-QTOF-HRMS. We note that the PFAS dosing experiment was designed and performed in 2019, prior to the knowledge of GenX degradation in acetone or acetonitrile reported by Liberatore et al. (2020). Therefore, quantitative analysis of GenX was compromised and not reported in this study.



**Figure 1.** Schematic illustrating three stages of our experimental design to investigate bioaccumulation of pesticides and PFAS in BSFL.

### 2.3.2 Feedstock source and characterization

Six batches of hulls, from nonpareil, pollinator, or Monterey almond varieties, were obtained for this study from Chico and Buttonwillow cities in California, USA (Figure 1). Hulls as received contained approximately 18% water and were ground using a hammer mill with a 6.35 mm screen. All hulls were stored air-tight at room temperature (up to 36 months). Optimization of BSFL growth on the same batch of hulls was conducted by Palma et al. (2020).

The method for extraction of TrOCs from hulls was modified from a protocol for extraction of dust developed by Moschet et al. (Moschet et al., 2018b) to achieve comprehensive recovery of TrOCs with a range of physiochemical properties using both liquid and gas chromatography HRMS. Briefly, 10 g of hulls were freeze dried (HarvestRight, UT, USA) for 16 hours (12 hr of drying cycle) and further ground using a Retsch MM 403 mixer mill (Verder Scientific, PA, USA) at a frequency of 30 Hz for 2 minutes. Each ground sample was passed through 106  $\mu\text{m}$  sieve, and 0.2 g of sieved sample was transferred into a 15 mL centrifuge tube. The samples were vortexed for 1 min in 3 mL of HPLC-grade hexane:acetone in 3-to-1 volume ratio (Fisher Scientific, MA, USA) and sonicated (Bransonic 8510E-MT, CT, USA) for 15 min. Samples were centrifuged at 3,500 rpm for 10 min, and all the supernatants were transferred to glass test tubes. The hull samples remaining in the centrifuge tubes were re-extracted with 3 mL of 100% acetone following the same procedure. The combined extracts (hexane:acetone and 100% acetone) were evaporated to ~1 mL under gentle  $\text{N}_2$  gas at 35 °C using a TurboVap (Biotage, UU, Sweden). Because the hulls contained higher lipid contents than dust, we used Captiva Enhanced Matrix Removal (EMR) cartridges (Agilent Technologies, CA, USA) for sample clean-up following the 5991-8308EN protocol (for 6 mL cartridge) to avoid contaminating the instruments. The extracts were solvent exchanged to Optima™ HPLC grade acetonitrile (Fisher Scientific, Pittsburgh, PA, USA). Excess (~2 g) magnesium sulfate anhydrous (Sigma-Aldrich, MO, USA) was used to remove any residual

water, and all extracts were passed through regenerated cellulose membranes (Captive Agilent Premium Syringe Filter, 0.2  $\mu\text{m}$ ) to remove any particles. Finally, we added an internal standard (ISTD) mixture containing 12.5 ng of each of three compounds—4,4'-dibromooctafluorobiphenyl (Sigma-Aldrich, MO, USA) for GC-QTOF-MS in EI and NCI ionization modes, imidacloprid-D<sub>4</sub> (Supelco INC, PA, USA) for LC positive ionization mode, and sodium perfluoro-[<sup>13</sup>C<sub>8</sub>]octanesulfonate (Wellington Laboratories, ON, CA) for LC negative ionization mode. Samples with ISTD were further evaporated to 0.25 mL for final analysis. Details of analytical standards of PFAS and pyrethroids are available in Appendix A (A1).

### 2.3.3 Larvae bioaccumulation reactors

We adopted the bioreactor design developed by Palma et al. (2018) for controlled larvae growth experiments. Each reactor contained 200 g of dry hull mass amended with distilled water to achieve 85% initial fiber saturation (69 $\pm$ 1% moisture content). The hulls at the end of incubation had a slightly higher water content (72 $\pm$ 2% moisture content). We used a flow meter to maintain the aeration rate at 40 mL/min, and the incubator temperature was set at 21°C. We adjusted the C/N ratio to 26 using urea (Fisher Scientific Company LLC, NH, USA) as a nitrogen source. In Stage 3 experiments, the feedstocks were first equilibrated for 1 hr with or without spiked chemicals (five-PFAS compounds) prior to the larvae growth experiments. The hulls were then incubated with or without 5-day-old BSFL larvae for 14 days (Figure 1). The initial BSFL were reared from eggs purchased from Symptom Black Soldier Fly (College Station, TX, USA). The detailed rearing method was described in Palma et al. (2018). In addition to the chemical analysis, samples of larvae were manually separated from the feed and weighed at harvest. Larvae from four reactors in Stage 3 were depurated prior to chemical analysis (see Figure 1), as detailed in A3. The end-of-experiment larval survival rates for all reactors were greater than 90% (Table A1 and Figure A2).

### 2.3.4 Larvae tissue extraction

We used the “quick, easy, cheap, effective, rugged, and safe” (QuEChERS) method to extract TrOCs from BSFL tissues. Approximately 5 g of ground larvae and 10 mL HPLC grade

acetonitrile (Fisher Scientific, MA, USA) were added to 50 mL extraction tubes and vortexed for 1 minute. QuEChERS salts packets (EN 15662) were added to each tube and vortex for another minute. The mixture was centrifuged (3,250 rpm) for 5 minutes and the supernatants evaporated to ~1 mL using N<sub>2</sub>. The larval extracts were cleaned up with EMR cartridges and spiked with the same type and quantity of ISTD mixture as the hull extracts. Finally, samples were evaporated to 0.25 mL before instrument analysis.

### *2.3.5 HPLC-MS analysis*

HPLC-MS was conducted on an Agilent 1260 Infinity HPLC system paired with a 6530 QTOF MS. Sample extracts (10 µL) were injected onto a C18 column (ZORBAX RRHD Eclipse Plus C18 column; 2.1 x 150 mm, 1.8 µm, Agilent Technologies, Inc.) at a flow rate of 0.4 mL/min, with a total run time of 31.5 min. The mobile phases for both ESI+ and ESI- modes were (A) 20 mM ammonium acetate (Fisher Scientific, Pittsburgh, PA, USA) in Optima™ HPLC grade water (Fisher Scientific, Pittsburgh, PA, USA) (A) and (B) acetonitrile. We used All-ion (data-independent) for data acquisition with collision energies (CE) of 0, 10, 20, and 40 eV, where 0 eV was used to acquire precursor ion information and other CE channels were used to acquire fragment ion information. Detailed data acquisition settings are available in Table A2.

### *2.3.6 GC-MS analysis*

GC-MS analysis was conducted on an Agilent 7890B gas chromatography paired with a 7200B QTOF MS using a HP-5MS 30 m × 0.25 mm, 0.25 µm column. In Stage 1, we ran sample extracts in EI and NCI modes following the acquisition method optimized by Moschet et al. (2017) for over 5,000 water contaminants (Moschet et al., 2017). In Stage 2 and 3, we only ran sample in NCI mode as this was much more sensitive for pyrethroid detection. The detailed instrumental parameters are available in Table A3 and A4.

### *2.3.7 Spike recovery test*

Spike recovery calculations were performed to sample matrices (Hull #5 and Tasty Grubs™ Larvae purchased from Tasty Worms, GA, USA). Percent recovery was calculated by taking the difference in concentration between pre-spiked samples (200 ng of pyrethroid and 24-PFAS standards were spiked before sample extraction) and non-spiked samples and dividing by the difference in concentration between post-spiked samples (200 ng of the same standard mix was spiked before the final evaporation step) and non-spiked samples (n=3 per matrix). Recoveries of quantified chemicals was 73-136% from hulls and 50-98% in BSFL (Figure A3).

### *2.3.8 Data processing and analysis*

We processed all MS data using a suspect-screening workflow we previously developed (W. Li et al., 2021). Briefly, the instrument-specific software—Agilent MassHunter Quantitative Analysis (B.09.00)—was used for target compound quantification. Agilent MassHunter Qualitative Analysis (B.08.00) was used for suspect-screening by applying the “Find by Formula” search against a combined personal compound database Library (PCDL), including Agilent GC-QTOF Pesticide PCDL (852 compounds with MS/MS spectra), the Water Contaminants PCDL (1,083 compounds with MS/MS spectra), and an in-house PFAS library (3,793 PFAS, 63 MS/MS spectra). Detailed suspect-screening criteria is listed in Table S5.

Nontargeted analysis was conducted on HPLC (ESI-) datasets using open-source software. We used Reifycs Abf Converter (<https://www.reifycs.com/AbfConverter/>; downloaded on June 9<sup>th</sup>, 2020) to convert raw data to analysis base file. Then we extracted and aligned chemical features of all data using MS-DIAL (v4.24).(Tsugawa et al., 2015) We used the retention times of ISTD for retention time correction. Finally, we used EnviHomolog (<http://www.envihomolog.eawag.ch>; accessed on February 8, 2021) to identify novel PFAS

homologous series that consist of repeating mass-increments of 49.9974 (-CF<sub>2</sub>-) and 99.9942 (-C<sub>2</sub>F<sub>4</sub>-).

## 2.4 Results and discussion

### 2.4.1 Qualification of residual pesticides and PFAS in almond hulls

Across the six types of almond hulls, we tentatively identified constituent structures in samples with level 3 confidence as outlined by Schymanski et al. (2014). We identified five compounds from LC (ESI-), 10 from LC (ESI+), 13 from GC (EI), and 18 from GC (NCI). To contextualize the level of concern that the identified compounds pose towards the ultimate consumers of BSFL intended for use as feed and food additives (e.g., livestock and humans), we categorized the identified chemicals as bioaccumulative (B), persistent (P), and/or toxic (T). We followed the categorization scheme outlined in the “Guidance on Information Requirements and Chemical Safety Assessment for Bioaccumulative Persistent and Toxic or Very Persistent and Very Bioaccumulative (BPT/vPvB) assessment” (ECHA, 2017). The qualified pesticides along with their ion abundance normalized by the ion abundance of the corresponding ISTD are presented as a Venn diagram in Figure 2 and Figure A4. Out of 46 qualified contaminants, we classified 29 qualified compounds as potentially bioaccumulative, 24 as potentially toxic, and 27 as potentially persistent. One insecticide, isophenos-methyl was classified as low-concern.

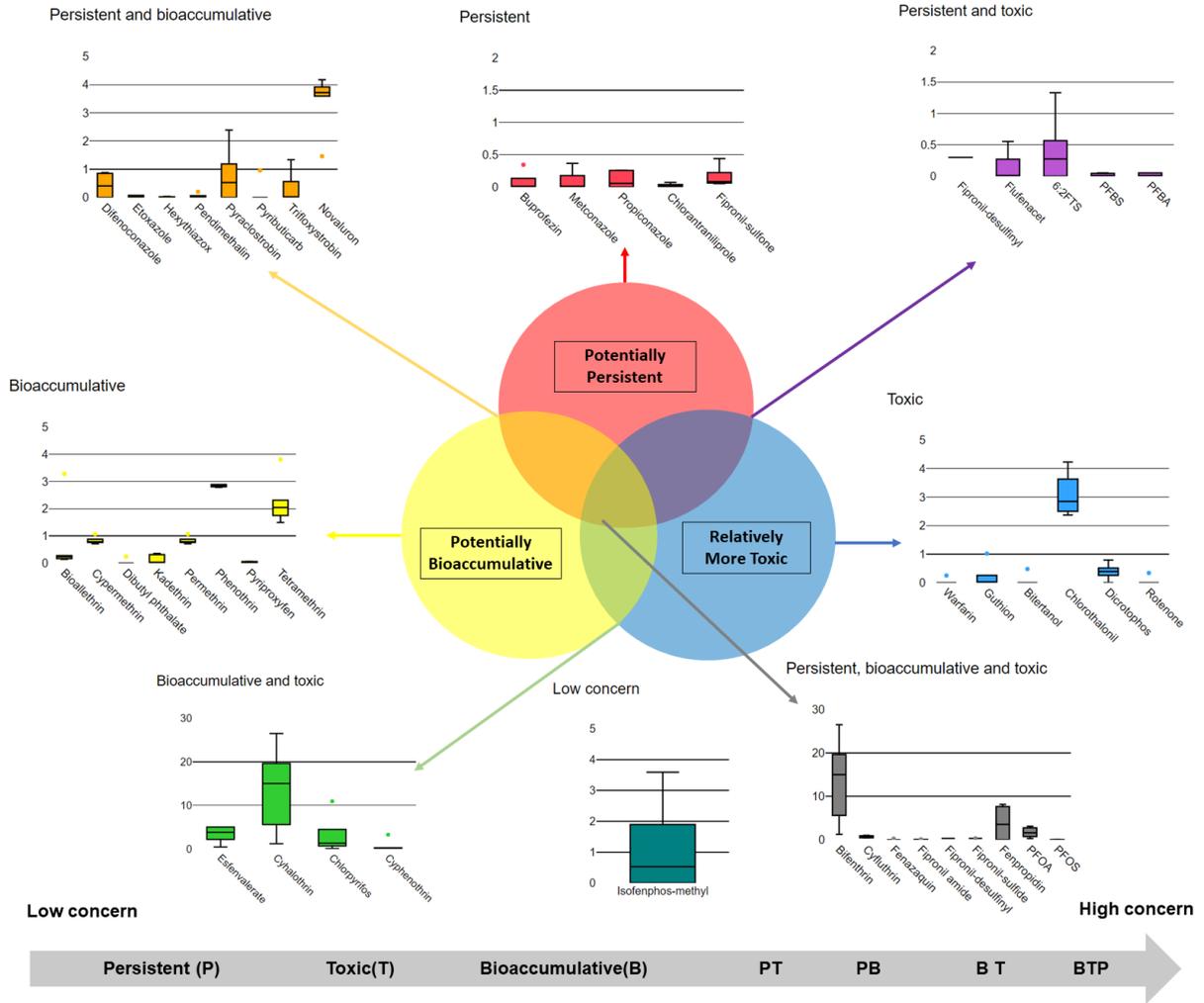
In the context of BSFL production from organic wastes and agricultural byproducts, we are interested in the relative level of concern that residual compounds pose for livestock (or even humans) via BSFL consumption. The guidelines stipulate that compounds are characterized as potentially bioaccumulative if the octanol-water partition coefficient ( $K_{ow}$ ) is over  $10^{4.5}$  (KOWWIN v1.67 estimate). Compounds are considered potentially persistent chemicals if 1) the

non-linear predicted biodegradation probability was less than 0.5 (BIOWIN 2), and 2) the ultimate biodegradation timeframe probability (BIOWIN 3) was less than 3. In other words, substances are considered persistent when the probability of biodegradation of the compounds within 3 months was less than 50%. To categorize relative toxicity, we prioritized mammalian (e.g., human, mouse, and rat) toxicity data rather than ecotoxicity data. Compounds with acute toxicity ( $<0.1$  mg/L or  $<0.1$  mg/kg) and chronic toxicity levels ( $<0.01$  mg/L or  $<0.01$  mg/kg) for reference dosage are classified as potentially toxic. For compounds that did not have mammalian toxicity reference dosage (fipronil amide and fipronil sulfide), we used available toxicity data and applied the same toxicity classification cut-off levels. For instance, we classified fipronil amide as a toxic compound based on its median lethal concentration (LC50) 0.0432 mg/L for yellow fever mosquito ( $<0.1$  mg/L). Detailed toxicity data and references used in this study are available in Tables A7.

We utilized the PBT classification scheme to select two hulls (Hull #3 and Hull #5) to use as substrate for downstream larvae bioaccumulation assessments (Stage 2 experiments in Figure 1). Our selection of the hulls was based on abundance of compounds of high concern (BPT). Interestingly, the most bioaccumulative compounds (categorized as BPT, BT, and BP) detected were also relatively more abundant in the hulls. We also narrowed the targeted quantification approach in the Stage 2 bioaccumulation experiment to a single analytical method—GC-QTOF-MS(NCI)—which is appropriate for quantification of pyrethroids (e.g., cyhalothrin, bifenthrin, etc).

The classification approach that we applied herein provided a framework to categorize and select organic waste substrates based on their potential level of concern. We recognize several limitations to the approach. First, use of Log  $K_{ow}$  for classification prioritizes lipid-associated bioaccumulation pathways, and could underestimate protein-associated bioaccumulation routes

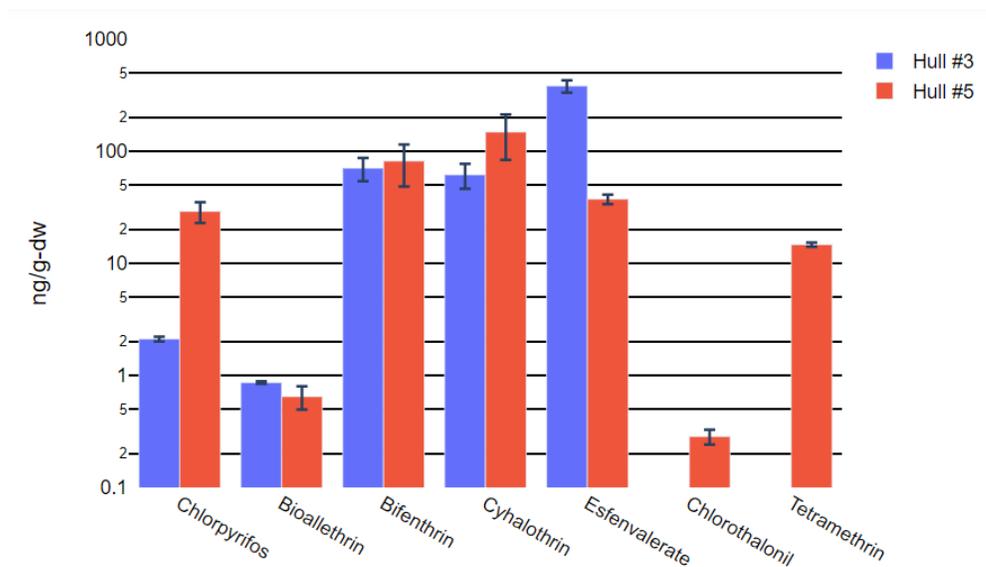
(Gottardo & Sokull-klüttgen, 2014). For instance, PFBS is not classified as bioaccumulative in Figure 2 despite strong binding to biomacromolecules (Allendorf et al., 2019; H. N. Bischel et al., 2011a; W. Li et al., 2021) and observed bioaccumulation in aquatic species including gastropods, crab, fish, and shrimp (L. Chen et al., 2018; Hong et al., 2015; P. Wang et al., 2020). Second, we used the biodegradation probability program (BIOWIN) to estimate biodegradability of all qualified compounds rather than using experimental biodegradation parameters. We generally consider experimental biodegradation parameters to be preferable to biodegradation estimates. However, three types of biodegradability tests are typical (i.e., ready biodegradability, inherent biodegradability, and biodegradation simulation tests), and a test could take place in diverse environmental compartments (e.g., water, marine, sediment, and soil). According to the guidelines, a substance is considered P/vP if it satisfies one of the three tests in any environmental compartment (ECHA, 2017). We used BIOWIN (Pavan & Worth, 2008) predicted values to streamline the classifications and to maintain consistency of data sources. Third, experimental data on toxicity for humans and livestock (i.e., the ultimate consumers of BSFL) was unavailable for fipronil amide and fipronil sulfide, so ecotoxicity data was used in place of mammalian toxicity values. Several reviews discuss challenges in extrapolating ecological toxicity to human and animal endpoints (Perkins et al., 2013; van den Berg et al., 2021; Q. Zhang et al., 2018).



**Figure 2.** Persistent, bioaccumulative, and relatively more toxic compounds identified in six almond hulls. The y-axis shows the ion abundance normalized by corresponding internal standards relevant for each analytical pipeline applied in this study.

### 2.4.2 Quantification of pyrethroids in almond hulls

Pyrethroid insecticide concentrations in the two hulls selected for stage 2 experiments ranged from 0.1 to 382 ng/g-dry hull (Figure 3). Five pyrethroids were quantified in Hull #3 and an additional two pyrethroids were detected in Hull #5. Recovery rates for the extraction of pesticides from almond hulls ranged from 81.8% to 128.7% (Figure S3). The most abundant pesticide in Hull #3 was esfenvalerate (381.6±47.7 ng/g-dry hull). The most abundant pesticides in Hull #5 were cyhalothrin (148.2±64.4 ng/g-dry hull) and bifenthrin (81.7±33.3 ng/g-dry hull). According to the Annual Statewide Pesticide Use Report (PUR) database, and consistent with our measurements, esfenvalerate, cyhalothrin, and bifenthrin were widely applied in almond orchards in the counties from which Hulls #3 and #5 were sourced (Table A6). These three pyrethroids were also shown to remain detectable level on almond tree twigs and barks in the field for over six months post application (Hamby et al., 2013).



**Figure 3.** Concentrations of pyrethroid pesticides quantified in Hulls #3 and #5. Almond hulls were used as substrates for investigating pyrethroid bioaccumulation in black soldier fly larvae (BSFL).

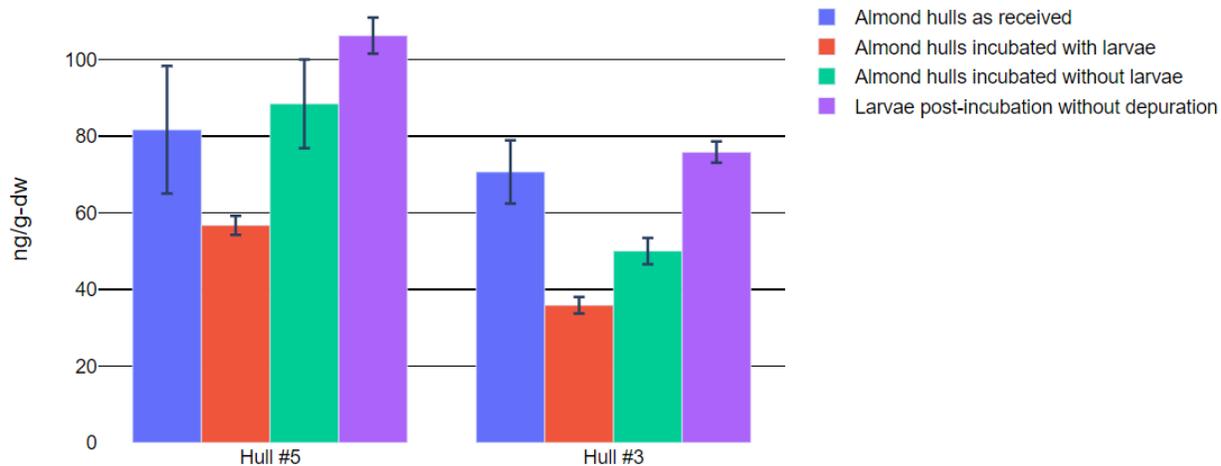
### 2.4.3 Bioaccumulation of bifenthrin in BSFL

Among all residual pyrethroids quantified in Hull #3 and Hull #5 in Stage 2 experiments, only bifenthrin bioaccumulated in BSFL. BSFL produced from Hull #3 contained  $75.8 \pm 2.8$  ng/g-dw bifenthrin, which was  $52.7 \pm 31.9\%$  higher than the level of bifenthrin in the residual hulls (Figure 4). Similarly, BSFL produced from Hull #5 contained  $46.6 \pm 29.5\%$  higher bifenthrin levels than the residual hulls. The level of bifenthrin in Hull #3 substrate significantly decreased after digestion by BSFL. Bifenthrin has the highest octanol-water partitioning coefficient ( $\text{Log } K_{ow} 8.15$ ) among all the residual pyrethroids detected in Hull #3 and Hull #5 (Table A7). Considering the high lipid content of BSFL ( $25.8 \pm 4\%$  of BSFL dry mass), we expected bioaccumulation of bifenthrin in larvae tissue relative to the substrates (Jackson et al., 2017). In comparison to BSFL, the hulls used have much lower lipid content (2.48% of Hull #3 dry mass and 2.65% of Hull #5 dry mass as previously determined by Palma et al. (2020). Decreases of bifenthrin in the substrate during BSFL digestion could be a result of bioaccumulation and/or biotransformation. Anaerobic bacteria in BSFL guts are known to degrade or promote degradation of a wide spectrum of organic chemicals, including some pharmaceuticals (Gold, Marie, et al., 2020a; Gold, von Allmen, et al., 2020b) Further research is needed to investigate the potential for BSFL-accelerated degradation of bifenthrin and the contribution of substrate microbial communities to bifenthrin degradation.

To assess whether bifenthrin accumulated in BSFL tissue (versus larval gut content), we performed a separate batch experiment during experimental Stage 3 (Figure 1). BSFL were raised on Hull #5, which was also spiked with five-PFAS compounds, and the bifenthrin concentrations were quantified in both non-depurated and depurated BSFL. The depuration process is detailed in Section A3. When larval gut content ( $21 \pm 2\%$  of their dry body weight) was

expelled from the larvae prior to analysis, the concentration of bifenthrin in larvae increased. The level of bifenthrin in depurated larvae (four replicate reactors) was  $44\pm 19\%$  higher than the level in non-depurated (four replicate reactors) larvae. These observations suggest that bifenthrin detected in BSFL samples was associated to a greater extent with larval tissues than with the residual gut content, assuming that the low level of added PFAS did not promote bifenthrin uptake in BSFL tissue. Typical practice in the cultivation and usage of BSFL does not include larvae depuration (e.g., blanching larvae in boiling water) to clear gut contents (Dortmans et al., 2017).

BSFL reared from feed streams containing known toxic, bioaccumulative, and/or persistent contaminants generally are used for non-consumptive uses, such as soil amendment, biofuel generation, or other non-consumptive commercially valuable products in the United States (Fernandez-Bayo et al., 2020; Y.-S. Wang & Shelomi, 2017). However, alternative post-production processes could enable safe use of BSFL with low levels of undesirable impurities as animal feed. For BSFL substrates that contain high levels of high-log- $K_{ow}$  TrOCs, separating the larvae into lipid and protein fractions could be beneficial for commercial applications and for managing detected bioaccumulative chemicals. TrOCs (e.g., bifenthrin) that primarily accumulate in lipid or fatty fractions of BSFL tissues may be eliminated from protein-based BSFL feeds by using defatting processes. The separated fatty fractions could be used in non-consumptive pipelines such as biofuel production (Manzano-agugliaro et al., 2012; Su et al., 2019; Y.-S. Wang & Shelomi, 2017), while improving the quality of the protein-rich end-products derived from BSFL tissues.



**Figure 4.** Bifenthrin concentrations measured in two almond hull types used for cultivation of black soldier fly larvae (BSFL), residual hulls following 14-days incubation with and without BSFL, and in the final BSFL.

#### 2.4.4 Uptake of spiked PFAS targets in BSFL

In Stage 3 experiments (Figure 1), we evaluated the potential bioaccumulation of PFAS in BSFL using spiked hulls and targeted HPLC-QTOF-MS analysis. Residual PFAS in all hulls from Stage 1 screenings were low, so targeted PFAS were spiked into Hull #5 (10 ppb final concentration of each of five PFAS—PFBA, PFOA, PFBS, PFOS, and GenX) for the experiment. While several spiked PFAS were detected in BSFL tissues following exposure, PFAS were not considered bioaccumulative ( $BAF < 1$ ) in BSFL (Figure A5). The most abundant PFAS detected in BSFL was PFOS, which was  $88 \pm 9.0\%$  of the levels in the substrate after digestion. The PFOS concentration further decreased to  $47 \pm 7.8\%$  of substrate levels after BSFL depurated for 24 hr. The concentration of PFOA in BSFL (without depuration) was  $70 \pm 10.3\%$  of the residual hulls, which further decreased to  $24 \pm 8.0\%$  after depuration. The end concentrations of PFBA and PFBS in BSFL (without depuration) were  $< 9\%$  of the residual hull, which further decreased to below the limit of quantification (LOQ). GenX was not quantifiable using the current extraction method. The spike recoveries of the extraction methods of the four spiked

PFAS for both hulls and BSFL were between 70% and 130% (Figure A3). Spiked PFAS levels did not negatively impact BSFL growth (Figure A2) or survival rates (Table A1.1 and A1.2). We observed that the bioaccumulation factor (BAF) of PFOS was higher than PFOA, which was consistent with the findings for earthworms detailed by Rich et al. (2015).

#### *2.4.5 Nontarget analysis of PFAS in almond hulls*

In Stage 1 experiments, we screened all-ion (ESI-) data against a large in-house PFAS library. While this screening method was extensive for analysis of TrOCs contamination on hulls, the method overlooks untargeted PFAS (e.g., compounds that are not included in the PFAS library). In Stage 3 experiments, we also collected all-ion data and subsequently extracted data from 10,444 unique chemical features. We screened the data using Kendrick mass defect (KMD) analysis and assigned homologous series using a previously established workflow (Myers et al., 2014; Y. Liu et al., 2019a; Bugsel & Zwiener, 2020; Koelmel et al., 2020; W. Li et al., 2021). KMD analysis utilizes the differences between exact mass and nominal mass. PFAS with the same functional group but different numbers of repeating mass units (e.g.  $-CF_2-$ , and  $-C_2F_2-$ ) have similar, if not the same, KMD (Barzen-Hanson et al., 2017). We performed KMD analysis and assigned homologous series for data obtained from (1) Hull #5 as received (not spiked with PFAS standards), (2) Hull #5 pre-spiked with 24 PFAS standards (PFAC-24PAR from Wellington Laboratories Inc., ON, CA), and (3) the 24-PFAS standard mixture alone. The pre-spiked hulls (used for extraction recovery test) showed distinctive classification patterns for perfluoroalkyl carboxylic acids (PFCAs), perfluoroalkyl sulfonic acids (PFSAs), and fluorotelomer sulfonates (FTSs) (Figure 5a). We did not detect perfluorooctane sulfonamido acetic acids (FOSAAs; two compounds) or perfluorooctane sulfonamide (FOSAs; one compound) using this method, due to lack of three repeating  $-CF_2-$  increments. Interestingly, we

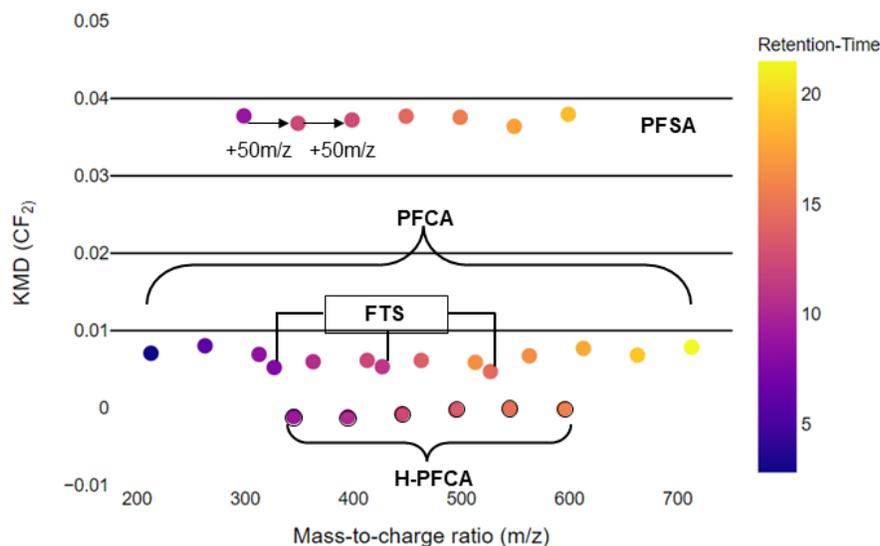
identified hydrogen-substituted polyfluorinated carboxylic acid homologous series (H-PFCA). H-PFCAs were also detected in Hull #5 as received (Figure A6a), but they were not present in the PFAS standards mix used to spike the hulls for recovery test (Figure A6b).

#### *2.4.6 Nontarget analysis reveals uptake of novel PFAS in BSFL*

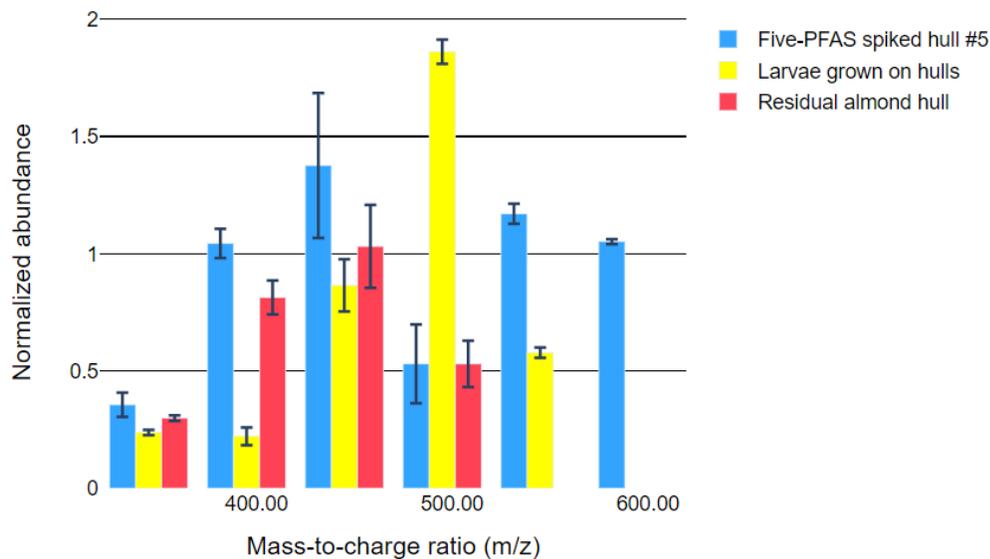
In Stage 3 experiments, we applied the same KMD workflow to BSFL cultivated on Hull #5 that was spiked with five-PFAS standards. The normalized peak area of H-PFCA in larvae versus five-PFAS spiked Hull #5 post-digestion increased as the perfluorocarbon chain increased up to 10H-perfluorodecanoic acid (H-PFDA) and decreased as the perfluorocarbon chain further elongated (Figure 5b). This bioaccumulative “reverse V shape” trend is consistent with PFCA protein binding trends for albumin proteins and peroxisome proliferator activated receptors (Ng & Hungerbühler, 2014; Khazaei et al., 2021; W. Li et al., 2021). To our knowledge, hydrogen-substituted polyfluorinated carboxylic acids (H-PFCAs) have not been detected in biological tissues. However, hydrogen-substituted polyfluorinated sulfonic acids (H-PFSAs) were detected and enriched in mouse serum in an Aqueous Film-Forming Foam (AFFF) dosing study (McDonough et al., 2020). In fact, we identified at least four (H-PFCA) across all the samples of BSFL produced from hulls after re-processing the previously acquired data. Figure A7 shows an example of the manual structural assignment of the mass spectra. We purchased 9H-Perfluorononanoic acid (H-PFNA) and 8H-Perfluorooctanoic acid (H-PFOA) standards and acquired their mass spectra for confirmation. Similar fragmentation features were observed in the standards and samples (Figure A8). Two H-PFCAs (H-PFOA and H-PFNA) were also detected in the blanks but the levels of ion abundance were much lower than the ion abundance of the same compounds in the BSFL samples (<10%).

The exact source of H-PFCA is unclear as we found limited information on this class of compounds. We suspect that fluorinated high-density polyethylene (HDPE) containers contributed to low levels of H-PFCA contamination of the pesticide product. EPA found eight PFCAs with 4 to 11 carbons in a laboratory test of pesticide products using targeted analysis (T. Nguyen, 2021). The analysis did not include H-PFCAs or other novel PFAS, but concluded that fluorinated polymers were produced during the plastic fluorination process, which led to the positive detection of PFAS in the pesticides. Washington et al. (2009) observed that acrylate-linked fluorotelomer and fluorotelomer-based polymers in soil could go through biodegradation, which produced H-PFOA, based on a plausible degradation pathway and mass spectral evidence (Washington et al., 2009; Washington, Jenkins, & Weber, 2015). Contamination of certain batches of mobile phase, or unavoidable fluorotelomer coated compartments in the pumps of LC could also contribute the positive detection of the homologous series as reported by Awchi et al. (2022). However, contamination was unlikely to be the cause of H-PFCAs detection in the larvae and hulls. H-PFCAs were highly abundant in BSFL, present with 100% frequency in Hull #5 samples (whether or not spiked with several PFAS standards) and were not detected in the analytical standards. Many researchers and policy makers have raised safety concerns of fluorotelomers production, uses, and recycling processes, due to the lack of end-of-life treatment, high persistence, and high likelihood of human exposures (Lohmann et al., 2020; Washington, Jenkins, Rankin, et al., 2015). From this study, we support this precautionary approach for managing fluorotelomers as bioaccumulative; non-target PFAS may be leaking from fluorotelomers and entering foods of humans.

a.



b.



**Figure 5.** a. Kendrick mass defect (KMD) analysis plots of Hull #5 that was spiked with 24 PFAS standards. The samples were used to determine percent recovery of PFAS from the extraction method. The color of each dot represents its retention time. H-PFCAs (labeled with dark circles) were not present in PFAS standards that were analyzed separately from the almond hulls. b. PFAS detected in Hull #5 and BSFL (depurated) that was cultivated on Hull #5. The size of each bar corresponds to the ion abundance normalized by the ion abundance of the internal standard (mass-labeled PFOS). Error bars represent the standard deviation of the normalized ion abundance of four replicates.

## 2.5 Conclusion

The suspect-screening approach we applied in this study offers a strategy for bioaccumulation assessments for BSFL cultivation that does not pre-select the chemicals of interest. Using this approach, we found that the bioaccumulation factor of bifenthrin for depurated larvae (14-day BAF =  $2.17 \pm 0.24$ ;) was higher than heavy metals that for (i.e. mercury 19-day BAF= $1.51 \pm 0.08$ ) determined by Proc et al. (2020). Our results are not surprising, as bifenthrin is classified as persistent, bioaccumulative, and toxic to mammals according to the *Guidance on Information Requirements and Chemical Safety Assessment*. We further built on the suspect-screening approach by applying KMD analysis of nontarget data acquired from almond hulls and BSFL. The nontarget analysis revealed the presence of an emerging class of PFAS (hydrogen-substituted polyfluoroalkyl carboxylic acids) in almond hulls and indicated that H-PFCAs accumulated in BSFL tissues with some dependency on the perfluorocarbon chain length. This study demonstrated a more comprehensive safety assessment pipeline for trace organic contaminants in nutrient recovery insects. The pipeline aimed to reduce bias associated with pre-selection of chemical targets of interest. Moving forward, we recommend application of more comprehensive chemical safety assessments when assessing novel feed and food products that pose unknown health risks.

## 2.6 References

- Allendorf, F., Berger, U., Goss, K. U., & Ulrich, N. (2019). Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. *Environmental Science: Processes and Impacts*, *21*(11), 1852–1863. <https://doi.org/10.1039/c9em00290a>
- 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. (2017). Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater. *Environmental Science and Technology*, *51*(4), 2047–2057. <https://doi.org/10.1021/acs.est.6b05843>
- Bhatt, P., Huang, Y., Zhan, H., & Chen, S. (2019). Insight into microbial applications for the biodegradation of pyrethroid insecticides. *Frontiers in Microbiology*, *10*(August), 1–19. <https://doi.org/10.3389/fmicb.2019.01778>
- Bischel, H. N., Macmanus-Spencer, L. A., Zhang, C., & Luthy, R. G. (2011). Strong associations of short-chain perfluoroalkyl acids with serum albumin and investigation of binding mechanisms. *Environmental Toxicology and Chemistry*, *30*(11). <https://doi.org/10.1002/etc.647>
- Bugsel, B., & Zwiener, C. (2020). LC-MS screening of poly- and perfluoroalkyl substances in contaminated soil by Kendrick mass analysis. *Analytical and Bioanalytical Chemistry*, *412*(20), 4797–4805. <https://doi.org/10.1007/s00216-019-02358-0>
- CALIFORNIA ALMOND INDUSTRY FACTS. (2016). Almond Board. [http://www.almonds.com/sites/default/files/2016\\_almond\\_industry\\_factsheet.pdf](http://www.almonds.com/sites/default/files/2016_almond_industry_factsheet.pdf)
- Chen, L., Tsui, M. M. P., Shi, Q., Hu, C., Wang, Q., Zhou, B., Lam, P. K. S., & Lam, J. C. W. (2018). Accumulation of perfluorobutane sulfonate (PFBS) and impairment of visual function in the eyes of marine medaka after a life-cycle exposure. *Aquatic Toxicology*, *201*(May), 1–10. <https://doi.org/10.1016/j.aquatox.2018.05.018>
- Costa, L. G. (2015). The neurotoxicity of organochlorine and pyrethroid pesticides. In *Handbook of Clinical Neurology* (1st ed., Vol. 131). Elsevier B.V. <https://doi.org/10.1016/B978-0-444-62627-1.00009-3>
- Dortmans, B., Diener, S., Verstappen, B., & Zurbrugg, C. (2017). *Black Soldier Fly Biowaste Processing*.
- ECHA. (2017). Guidance on Information Requirements and Chemical Safety Assessment (Chapter R.11: PBT/vPvB assessment, Version 3.0). In *EUROPEAN CHEMICALS AGENCY (ECHA)* (Issue June). <https://doi.org/10.2823/128621>
- Emr, A. C. (n.d.). *Agilent Captiva EMR-Lipid Method Guide*. 1–4.
- Evich, M. G., Davis, M. J. B., McCord, J. P., Acrey, B., Awkerman, J. A., Knappe, D. R. U., Lindstrom, A. B., Speth, T. F., Tebes-Stevens, C., Strynar, M. J., Wang, Z., Weber, E. J., Henderson, W. M., & Washington, J. W. (2022). Per- and polyfluoroalkyl substances in the environment. *Science*, *375*(6580). <https://doi.org/10.1126/science.abg9065>
- Fernandez-Bayo, J. D., Shea, E. A., Parr, A. E., Achmon, Y., Stapleton, J. J., VanderGheynst, J. S., Hodson, A. K., & Simmons, C. W. (2020). Almond processing residues as a source of organic acid biopesticides during biosolarization. *Waste Management*, *101*, 74–82. <https://doi.org/10.1016/j.wasman.2019.09.028>
- Filou, E. (2021). Bugs to the rescue: using insects as animal feed could cut deforestation – report. *The Guardian*, 1–3.

- Fowles, T. M., & Nansen, C. (2019). *Artificial selection of insects to bioconvert pre-consumer organic wastes . A review.*
- Glüge, J., Scheringer, M., Cousins, I. T., Dewitt, J. C., Goldenman, G., Herzke, D., Lohmann, R., Ng, C. A., Trier, X., & Wang, Z. (2020). An overview of the uses of per- And polyfluoroalkyl substances (PFAS). *Environmental Science: Processes and Impacts*, 22(12), 2345–2373. <https://doi.org/10.1039/d0em00291g>
- Gold, M., Marie, C., Zurbrügg, C., Kreuzer, M., Boulos, S., Diener, S., & Mathys, A. (2020). Biowaste treatment with black soldier fly larvae : Increasing performance through the formulation of biowastes based on protein and carbohydrates. *Waste Management*, 102, 319–329. <https://doi.org/10.1016/j.wasman.2019.10.036>
- Gold, M., Tomberlin, J. K., Diener, S., Zurbrügg, C., & Mathys, A. (2018). Decomposition of biowaste macronutrients, microbes, and chemicals in black soldier fly larval treatment: A review. *Waste Management*, 82, 302–318. <https://doi.org/10.1016/j.wasman.2018.10.022>
- Gold, M., von Allmen, F., Zurbrügg, C., Zhang, J., & Mathys, A. (2020). Identification of Bacteria in Two Food Waste Black Soldier Fly Larvae Rearing Residues. *Frontiers in Microbiology*, 11(November). <https://doi.org/10.3389/fmicb.2020.582867>
- Gottardo, S., & Sokull-klüttgen, B. (2014). *Review of available criteria for non-aquatic organisms within PBT / vPvB frameworks.* <https://doi.org/10.2788/11347>
- Hamby, K. A., Alifano, J. A., & Zalom, F. G. (2013). Total effects of contact and residual exposure of bifenthrin and  $\lambda$ -cyhalothrin on the predatory mite *Galendromus occidentalis* (Acari: Phytoseiidae). *Experimental and Applied Acarology*, 61(2), 183–193. <https://doi.org/10.1007/s10493-013-9680-z>
- Hart, J., Somerville, S., & Sumner, D. A. (2020). *Economic, Resource, and Environmental Consequences of Changes in By-Product use by Dairies. October*, 1–23.
- Hatzenbuehler, P. L., Tejada, H., Hines, S., & Packham, J. (2021). Change in Hay-To-Milk Price Responsiveness with Dairy Industry Expansion. *Journal of Agricultural and Applied Economics*, 53(2), 246–258. <https://doi.org/10.1017/aae.2021.3>
- Hensema, T. J., Berendsen, B. J. A., & van Leeuwen, S. P. J. (2021). Non-targeted identification of per- and polyfluoroalkyl substances at trace level in surface water using fragment ion flagging. *Chemosphere*, 265, 128599. <https://doi.org/10.1016/j.chemosphere.2020.128599>
- Hollender, J., Bavel, B. Van, Dulio, V., Farmen, E., Furtmann, K., & Koschorreck, J. (2019). High resolution mass spectrometry - based non - target screening can support regulatory environmental monitoring and chemicals management. *Environmental Sciences Europe*. <https://doi.org/10.1186/s12302-019-0225-x>
- Hong, S., Khim, J. S., Wang, T., Naile, J. E., Park, J., Kwon, B. O., Song, S. J., Ryu, J., Codling, G., Jones, P. D., Lu, Y., & Giesy, J. P. (2015). Bioaccumulation characteristics of perfluoroalkyl acids (PFAAs) in coastal organisms from the west coast of South Korea. *Chemosphere*, 129, 157–163. <https://doi.org/10.1016/j.chemosphere.2014.06.023>
- Id, N. M., Rijk, T. De, Loon, J. J. A. Van, & Zoet, L. (2021). *Effects of insecticides on mortality , growth and bioaccumulation in black soldier fly ( Hermetia illucens ) larvae.* 1–16. <https://doi.org/10.1371/journal.pone.0249362>
- Jackson, E., Shoemaker, R., Larian, N., & Cassis, L. (2017). Adipose tissue as a site of toxin accumulation. *Comprehensive Physiology*, 7(4), 1085–1135. <https://doi.org/10.1002/cphy.c160038>
- Kauffman, D. (2021). *Analytical Report For PEER (410-31526-1)* (Issue 850).

- Khazaei, M., Christie, E., Cheng, W., Michalsen, M., Field, J., & Ng, C. (2021). *Perfluoroalkyl Acid Binding with Peroxisome Proliferator - Activated Receptors  $\alpha$ ,  $\gamma$ , and  $\delta$ , and Fatty Acid Binding Proteins by Equilibrium Dialysis with a Comparison of Methods*. 1–16.
- Koelmel, J. P., Paige, M. K., Aristizabal-Henao, J. J., Robey, N. M., Nason, S. L., Stelben, P. J., Li, Y., Kroeger, N. M., Napolitano, M. P., Savvaides, T., Vasiliou, V., Rostkowski, P., Garrett, T. J., Lin, E., Deigl, C., Jobst, K., Townsend, T. G., Godri Pollitt, K. J., & Bowden, J. A. (2020). Toward Comprehensive Per- and Polyfluoroalkyl Substances Annotation Using FluoroMatch Software and Intelligent High-Resolution Tandem Mass Spectrometry Acquisition. *Analytical Chemistry*, 92(16), 11186–11194. <https://doi.org/10.1021/acs.analchem.0c01591>
- Kumar, S., Negi, S., Mandpe, A., Singh, R. V., & Hussain, A. (2018). Rapid composting techniques in Indian context and utilization of black soldier fly for enhanced decomposition of biodegradable wastes - A comprehensive review. *Journal of Environmental Management*, 227(May), 189–199. <https://doi.org/10.1016/j.jenvman.2018.08.096>
- Lara-Martín, P. A., Chiaia-Hernández, A. C., Biel-Maeso, M., Baena-Nogueras, R. M., & Hollender, J. (2020). Tracing Urban Wastewater Contaminants into the Atlantic Ocean by Nontarget Screening. *Environmental Science and Technology*, 54(7), 3996–4005. <https://doi.org/10.1021/acs.est.9b06114>
- Li, W., Hu, Y., & Bischel, H. N. (2021). In-vitro and in-silico assessment of per- and polyfluoroalkyl substances (PFAS) in aqueous film-forming foam (AFFF) binding to human serum albumin. *Toxics*, 9(3). <https://doi.org/10.3390/toxics9030063>
- Li, X., Giles, D. K., Niederholzer, F. J., Andaloro, J. T., Lang, E. B., & Watson, L. J. (2021). Evaluation of an unmanned aerial vehicle as a new method of pesticide application for almond crop protection. *Pest Management Science*, 77(1), 527–537. <https://doi.org/10.1002/ps.6052>
- Liberatore, H. K., Jackson, S. R., Strynar, M. J., & Mccord, J. P. (2020). Solvent Suitability for HFPO-DA (“GenX” Parent Acid) in Toxicological Studies. *Environmental Science and Technology Letters*, 7(7), 477–481. <https://doi.org/10.1021/acs.estlett.0c00323>
- Liu, Y., D’Agostino, L. A., Qu, G., Jiang, G., & Martin, J. W. (2019). High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly- and perfluoroalkyl substances (PFASs) in environmental and human samples. *TrAC - Trends in Analytical Chemistry*, 121, 115420. <https://doi.org/10.1016/j.trac.2019.02.021>
- Lohmann, R., Cousins, I. T., Dewitt, J. C., Glüge, J., Goldenman, G., Herzke, D., Lindstrom, A. B., Miller, M. F., Ng, C. A., Patton, S., Scheringer, M., Trier, X., & Wang, Z. (2020). Are Fluoropolymers Really of Low Concern for Human and Environmental Health and Separate from Other PFAS? *Environmental Science and Technology*, 54(20), 12820–12828. <https://doi.org/10.1021/acs.est.0c03244>
- Manzano-agugliaro, F., Sanchez-muros, M. J., Barroso, F. G., Martínez-sánchez, A., & Rojo, S. (2012). *Insects for biodiesel production*. 16, 3744–3753. <https://doi.org/10.1016/j.rser.2012.03.017>
- Mcdonough, C. A., Choyke, S., Ferguson, P. L., Dewitt, J., & Higgins, C. P. (2020). Bioaccumulation of Novel Per-and Polyfluoroalkyl Substances (PFASs) in Mice Dosed with an Aqueous Film-Forming Foam (AFFF). *Environmental Science & Technology*. <https://doi.org/10.1021/acs.est.0c00234>
- Miller, R. F., Baskett, R. S., & Scott, C. E. (n.d.). *Almond Hulls tested as feed for dairy cattle and lambs showed promise and limitations in value Red Kidney Beans symptoms of zinc deficiency disappear following application of foliage spray*.

- Moschet, C., Anumol, T., Lew, B. M., Bennett, D. H., & Young, T. M. (2018). Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. *Environmental Science and Technology*, *52*(5), 2878–2887. <https://doi.org/10.1021/acs.est.7b05767>
- Moschet, C., Lew, B. M., Hasenbein, S., Anumol, T., & Young, T. M. (2017). LC- and GC-QTOF-MS as Complementary Tools for a Comprehensive Micropollutant Analysis in Aquatic Systems. *Environmental Science and Technology*, *51*(3), 1553–1561. <https://doi.org/10.1021/acs.est.6b05352>
- Myers, A. L., Jobst, K. J., Mabury, S. A., & Reiner, E. J. (2014). Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products. *Journal of Mass Spectrometry*, *49*(4), 291–296. <https://doi.org/10.1002/jms.3340>
- Ng, C. A., & Hungerbühler, K. (2014). Bioaccumulation of perfluorinated alkyl acids: Observations and models. *Environmental Science and Technology*, *48*(9), 4637–4648. <https://doi.org/10.1021/es404008g>
- Nguyen, T. (2021). Rinses from Selected Fluorinated and Non-Fluorinated HDPE Containers. *United States Environmental Protection Agency*, 20460(Level 3), 1–10.
- Nguyen, T. T. X., Tomberlin, J. K., & Vanlaerhoven, S. (2015). Ability of Black Soldier Fly (Diptera: Stratiomyidae) Larvae to Recycle Food Waste. *Environmental Entomology*, *44*(2), 406–410. <https://doi.org/10.1093/ee/nvv002>
- Palma, L., Ceballos, S. J., Johnson, P. C., Niemeier, D., Pitesky, M., & VanderGheynst, J. S. (2018). Cultivation of black soldier fly larvae on almond byproducts: impacts of aeration and moisture on larvae growth and composition. *Journal of the Science of Food and Agriculture*, *98*(15), 5893–5900. <https://doi.org/10.1002/jsfa.9252>
- Palma, L., Fernandez-Bayo, J., Niemeier, D., Pitesky, M., & VanderGheynst, J. S. (2019). Managing high fiber food waste for the cultivation of black soldier fly larvae. *Npj Science of Food*, *3*(1). <https://doi.org/10.1038/s41538-019-0047-7>
- Pavan, M., & Worth, A. P. (2008). Review of estimation models for biodegradation. *QSAR and Combinatorial Science*, *27*(1), 32–40. <https://doi.org/10.1002/qsar.200710117>
- Perednia, D. A., Anderson, J., & Rice, A. (2017). A comparison of the greenhouse gas production of Black Soldier Fly Larvae versus aerobic microbial decomposition of an organic feed material. *Research & Reviews: Journal of Ecology and Environmental Sciences*, *5*(3), 10–16.
- Perkins, E. J., Ankley, G. T., Crofton, K. M., Garcia-reyero, N., Lalone, C. A., Johnson, M. S., Tietge, J. E., & Villeneuve, D. L. (2013). Alternative species in risk assessment Use of Alternative Species and in Vitro Assays in an AOP Framework. *Environmental Health Perspectives*, *121*(9), 1002–1010.
- Proc, K.; Bulak, P.; Wi, D.; Bieganski, A. Science of the Total Environment *Hermetia Illucens Exhibits Bioaccumulative Potential for 15 Different Elements – Implications for Feed and Food Production*. **2020**, 723. <https://doi.org/10.1016/j.scitotenv.2020.138125>.
- Prosekov, A. Y., & Ivanova, S. A. (2018). Food security: The challenge of the present. *Geoforum*, *91*(February), 73–77. <https://doi.org/10.1016/j.geoforum.2018.02.030>
- Purschke, B., Scheibelberger, R., Axmann, S., Adler, A., & Jäger, H. (2017). Impact of substrate contamination with mycotoxins, heavy metals and pesticides on the growth performance and composition of black soldier fly larvae (*Hermetia illucens*) for use in the feed and food value chain. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, *34*(8), 1410–1420. <https://doi.org/10.1080/19440049.2017.1299946>

- Rich, C. D., Blaine, A. C., Hundal, L., & Higgins, C. P. (2015). Bioaccumulation of perfluoroalkyl acids by earthworms (*Eisenia fetida*) exposed to contaminated soils. *Environmental Science and Technology*, 49(2), 881–888. <https://doi.org/10.1021/es504152d>
- Salomone, R., Saija, G., Mondello, G., Giannetto, A., Fasulo, S., & Savastano, D. (2017). Environmental impact of food waste bioconversion by insects: Application of Life Cycle Assessment to process using *Hermetia illucens*. *Journal of Cleaner Production*, 140, 890–905. <https://doi.org/10.1016/j.jclepro.2016.06.154>
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology*, 48(4), 2097–2098. <https://doi.org/10.1021/es5002105>
- Shumo, M., Osuga, I. M., Khamis, F. M., Tanga, C. M., Fiaboe, K. K. M., Subramanian, S., Ekesi, S., van Huis, A., & Borgemeister, C. (2019). The nutritive value of black soldier fly larvae reared on common organic waste streams in Kenya. *Scientific Reports*, 9(1), 1–13. <https://doi.org/10.1038/s41598-019-46603-z>
- Soldier, B., Larvae, F. F., Is, M., Profitable, M., & Meal, F. (2021). *Black Soldier Fly Full-Fat Larvae Meal Is More Profitable Than Effects on Aquaculture Sustainability , Economy and Fish*. 1–13.
- Su, C. H., Nguyen, H. C., Bui, T. L., & Huang, D. L. (2019). Enzyme-assisted extraction of insect fat for biodiesel production. *Journal of Cleaner Production*, 223, 436–444. <https://doi.org/10.1016/j.jclepro.2019.03.150>
- Surendra, K. C., Olivier, R., Tomberlin, J. K., Jha, R., & Kumar, S. (2016). Bioconversion of organic wastes into biodiesel and animal feed via insect farming. *Renewable Energy*, 98, 197–202. <https://doi.org/10.1016/j.renene.2016.03.022>
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., Vandergheynst, J., Fiehn, O., & Arita, M. (2015). *MS / MS deconvolution for comprehensive metabolome analysis*. 12(6). <https://doi.org/10.1038/nmeth.3393>
- US Environmental Protection Agency. (2021). *Analysis of PFAS in selected mosquito control products from the Maryland Department of Agriculture*. *ACB Project # B21-19*.
- van den Berg, S. J. P., Maltby, L., Sinclair, T., Liang, R., & van den Brink, P. J. (2021). Cross-species extrapolation of chemical sensitivity. *Science of the Total Environment*, 753, 141800. <https://doi.org/10.1016/j.scitotenv.2020.141800>
- Wang, P., Lu, Y., Su, H., Su, C., Johnson, A. C., Yu, L., & Jenkins, A. (2020). Managing health risks of perfluoroalkyl acids in aquatic food from a river-estuary-sea environment affected by fluorochemical industry. *Environment International*, 138(March), 105621. <https://doi.org/10.1016/j.envint.2020.105621>
- Wang, Y.-S., & Shelomi, M. (2017). Review of Black Soldier Fly (*Hermetia illucens*) as Animal Feed and Human Food. *Foods*, 6(91). <https://doi.org/10.3390/foods6100091>
- Washington, J. W., Ellington, J. J., Evans, J. J., Yoo, H., & Hafner, S. C. (2009). Degradability of an Acrylate-Linked , Fluorotelomer Polymer in Soil. *Environmental Science & Technology*, 43, 6617–6623.
- Washington, J. W., Jenkins, T. M., Rankin, K., & Naile, J. E. (2015). Decades-scale degradation of commercial, side-chain, fluorotelomer-based polymers in soils and water. *Environmental Science and Technology*, 49(2), 915–923. <https://doi.org/10.1021/es504347u>
- Washington, J. W., Jenkins, T. M., & Weber, E. J. (2015). Identification of Unsaturated and 2H Polyfluorocarboxylate Homologous Series and Their Detection in Environmental Samples and as Polymer Degradation Products. *Environmental Science and Technology*, 49(22), 13256–13263. <https://doi.org/10.1021/acs.est.5b03379>

- Williams, S. R. O., Chaves, A. V., Deighton, M. H., Jacobs, J. L., Hannah, M. C., Ribaux, B. E., Morris, G. L., Wales, W. J., & Moate, P. J. (2018). Influence of feeding supplements of almond hulls and ensiled citrus pulp on the milk production, milk composition, and methane emissions of dairy cows. *Journal of Dairy Science*, *101*(3), 2072–2083. <https://doi.org/10.3168/jds.2017-13440>
- Win, S. S., Ebner, J. H., Brownell, S. A., Pagano, S. S., Cruz-Diloné, P., & Trabold, T. A. (2018). Anaerobic digestion of black soldier fly larvae (BSFL) biomass as part of an integrated biorefinery. *Renewable Energy*, *127*, 705–712. <https://doi.org/10.1016/j.renene.2018.04.093>
- Zhan, Y., & Zhang, M. (2014). Science of the Total Environment Spatial and temporal patterns of pesticide use on California almonds and associated risks to the surrounding environment. *Science of the Total Environment*, *The*, *472*, 517–529. <https://doi.org/10.1016/j.scitotenv.2013.11.022>
- Zhang, Q., Li, J., Middleton, A., Bhattacharya, S., & Conolly, R. B. (2018). Bridging the Data Gap From in vitro Toxicity Testing to Chemical Safety Assessment Through Computational Modeling. *Frontiers in Public Health*, *6*(September), 1–16. <https://doi.org/10.3389/fpubh.2018.00261>
- Zhang, X., Liu, X., Luo, Y., & Zhang, M. (2008). Evaluation of water quality in an agricultural watershed as affected by almond pest management practices. *42*, 3685–3696. <https://doi.org/10.1016/j.watres.2008.05.018>
- Zheng, L., Hou, Y., Li, W., Yang, S., Li, Q., & Yu, Z. (2012). Biodiesel production from rice straw and restaurant waste employing black soldier fly assisted by microbes. *Energy*, *47*(1), 225–229. <https://doi.org/10.1016/j.energy.2012.09.006>

**CHAPTER 3: IN-VITRO AND IN-SILICO ASSESSMENT OF PER- AND  
POLYFLUOROALKYL SUBSTANCES (PFAS) IN AQUEOUS FILM-FORMING FOAM  
(AFFF) BINDING TO HUMAN SERUM ALBUMIN**

**Wenting Li <sup>1</sup>, Yuhong Hu <sup>1</sup> and Heather N. Bischel \***

Department of Civil and Environmental Engineering, University of California Davis, California,  
CA 95616, USA; lwenting@ucdavis.edu (W.L.); 3160101554@zju.edu.cn (Y.H.)

\*Correspondence: hbischel@ucdavis.edu; Tel.: +1 530-752-6772

Published - Toxics

St. Alban-Anlage 66 4052 Basel, Switzerland

<https://doi.org/10.3390/toxics9030063>

### 3.1 Abstract

Drinking water contaminated by fluorosurfactant-based aqueous film-forming foams (AFFF) is a source of human exposure to poly- and perfluoroalkyl substances (PFAS). However, assessment of bioaccumulation potentials of diverse PFAS in commercial products such as AFFF have been insufficient and challenging, especially due to a lack of analytical standards. Here we explore the value of suspect screening, equilibrium dialysis, and molecular-docking simulations to identify potentially bioaccumulative PFAS. We exposed human serum albumin (HSA) protein to dilutions of a legacy AFFF produced by 3M in 1999 using equilibrium dialysis and screened in-vitro protein-binding affinities using high-resolution mass spectrometry (HRMS). Through suspect screening, we identified 32 PFAS and 18 hydrocarbon surfactants in the AFFF that bound to HSA. Quantification of noncovalent association constants for 26 PFAS standards confirmed that many PFAS, including the short-chain perfluoropropane sulfonic acid ( $\log K_a = 4.1 \pm 0.2 \text{ M}^{-1}$ ), exhibit strong binding affinities with HSA. At least five PFAS in AFFF (including three PFAS with less than five perfluorocarbons) remained bound to the precipitated HSA pellet after extensive solvent washing—an indication of high PFAS binding potential. Three PFAS (PFBS, PFOS, and PFOA) were confirmed in the protein pellet with analytical standards and quantified after acid digestion—this sample fraction accounted for 5 to 20% of each compound mass in the sample. We calculated pseudo-bioconcentration factors ( $\text{BCF}_{\text{pseudo}}$ ) for PFAS that suspect screening flagged as noncovalently bound or potentially covalently bound. Most PFAS exhibiting high  $\text{BCF}_{\text{pseudo}}$ , especially those with seven perfluorocarbons, contained a carboxylic acid or a sulfonic acid. Finally, we used molecular docking to simulate HSA binding affinities for 62 ligands (26 PFAS targets, 18 PFAS qualified in AFFF, and 18 hydrocarbon surfactants qualified in AFFF). We found that molecular docking can effectively separate HSA-binding and -nonbinding compounds in

AFFF. In-vitro and in-silico approaches described in this study provide replicable, high-throughput workflows for assessing bioaccumulation potentials of diverse PFAS in commercial products.

### **3.2 Introduction**

Application of aqueous film-forming foams (AFFF) for fire-suppression at military bases and airports is a cause of drinking water contamination with poly- and perfluoroalkyl substances (PFAS) (Banzhaf et al., 2017). Fluorosurfactant-based legacy AFFF formulations include complex mixtures of perfluoroalkyl carboxylic acids (PFCA), perfluoroalkyl sulfonic acids (PFSA), and highly diverse PFAS, including polyfluorinated precursors (Barzen-Hanson et al., 2017). Consumption of AFFF-contaminated drinking water can lead to elevated PFAS levels in human blood (Gyllenhammar et al., 2015; Lingala & Ghany, 2016). Occupational exposure to PFAS in AFFF may also present health risks to firefighters (Daniels et al., 2015; Rotander et al., 2015; Trowbridge et al., 2020). Human exposure to PFAS has been linked to cancer, cardiovascular disease, kidney disease, liver disease, immune suppression, neurological disease, type II diabetes, osteoarthritis, respiratory disease, among other impacts (ATSDR, 2017; Lau et al., 2015).

Given these problems, researchers have gained interest in studying the health impacts of novel PFAS in AFFF (Sunderland et al., 2019; Z. Wang et al., 2017), including compounds with one perfluorinated carbon that sometimes are not classified as PFAS (e.g., fluorinated aromatics) (OECD, 2018a). In particular, constructing physiologically-based pharmacokinetic (PBPK) models for PFAS exposure requires researchers to determine affinities (quantified as partition coefficients or association constants) among different PFAS structures in different biological tissues. The high numbers and structural diversity of existing and emerging PFAS renders this task experimentally infeasible. An alternative approach is to evaluate associations of PFAS mixtures in

an AFFF with abundant model proteins (commonly serum albumin and liver fatty acid binding proteins) to identify potentially bioaccumulative PFAS and to yield quantitative relationships between PFAS exposure, bioaccumulation, and tissue distribution (Ng & Hungerbühler, 2014). In this study, we assess binding affinities of diverse PFAS in AFFF to human serum albumin (HSA). HSA is the most abundant protein in human blood plasma, presenting in tissues throughout the body, and serves important biological functions (e.g., transportation of fatty acids, drugs, and thyroid hormones) (Sobolewski et al., 2018). Data on the binding affinities of AFFF-derived PFAS with HSA will support development of PBPK models for such PFAS.

Accurately incorporating protein binding affinities into PBPK models requires accurate understanding and quantification of the molecular mechanisms at play. Several studies reported that PFCAs and PFSAAs bind with serum albumin proteins noncovalently through specific site binding or non-specific surface adsorption (Beesoon & Martin, 2015; H. N. H. N. Bischel et al., 2010; Woodcroft et al., 2010; Wu et al., 2009). Two studies identified a potential for covalent binding between PFAS and albumin proteins (Rand & Mabury, 2012; vanden Heuvel et al., 1992). While no studies have investigated ultrastrong noncovalent bindings ( $K_A > 10^9 \text{ M}^{-1}$ ) between PFAS and proteins, ultrastrong binding was observed for PFAS in aqueous supramolecular polymerization (Krieg et al., 2014). Common serum-extraction protocols are likely to overlook or discard strongly bound ligands, including covalently bound or ultrastrong noncovalently bound PFAS. In organic solvent extraction, for instance, the precipitated protein pellet is disposed after extraction—along with any strongly bound ligands or residual targets. In online or offline solid phase extraction (SPE), covalently bound ligands and denatured proteins are lost on SPE cartridges (Calafat et al., 2007). Additionally, many studies accounted for matrix effects by spiking calibration standards into blank serum before analysis (Pei et al., 2017). Extraction efficiencies are

typically reported as satisfactory (e.g., 70–130% spike recoveries) by using such matrix (serum) matched calibration curves for quantification, thereby inadvertently masking strong protein–ligand interactions (Allendorf et al., 2019; H. N. Bischel et al., 2011b).

**In this study, we combined experimental and modeling techniques to identify potentially bioaccumulative PFAS present in an AFFF and to investigate multiple binding pathways for diverse PFAS structures.** We first utilized HSA as a model protein system to quantify both noncovalently bound and potentially covalently bound PFAS using targeted analysis. This analytical approach facilitates evaluation of the degree to which strongly bound or residual PFAS may be discarded in precipitated protein pellets. We then estimated bioconcentration factors for legacy and novel PFAS using linear (L)-PFOS as a bioaccumulative benchmarking compound. Finally, we predicted protein binding affinities for novel PFAS using molecular docking, a traditional drug design tool that simulates interactions between small molecules and large proteins (Ng & Hungerbuehler, 2015; L. Zhang et al., 2013). The results comprised the most comprehensive quantification of relative PFAS-HSA binding affinities to date, providing valuable inputs for bioaccumulation and PBPK models.

### **3.3 Method and material**

#### *3.3.1 Study Design and Workflow*

The overall workflow consists of the following components. First, we exposed HSA to either (a) dilutions of an AFFF produced by 3M in 1999 or (b) in-house mixtures of 26 PFAS (listed in Table B6.2) through equilibrium dialysis. We used data-dependent (tMS/MS, for target PFAS) and data-independent (all-ion fragmentation, for suspect-screening) mass spectrometry against an in-house library of PFAS and hydrocarbon surfactants to identify compounds that were

bound noncovalently to HSA. Second, we acid-digested residual protein pellets that were free of noncovalently bound PFAS. We applied suspect screening to identify residual PFAS in the precipitated protein pellet. Residual PFAS were considered candidates for forming ultrastrong or covalent associations with HSA. We used targeted MS/MS for three PFAS to quantify residual levels in the protein pellet. Third, we quantitatively evaluated protein association constants predicted by molecular docking between 26 target PFAS structures and two HSA crystal structures. Fourth and finally, we used the molecular docking workflow to classify PFAS and non-PFAS surfactants in the AFFF as either HSA-binding or non-binding.

### 3.3.2 *Equilibrium Dialysis*

Equilibrium dialysis was performed in a 96-well system (Harvard Apparatus) in which each polypropylene cell was separated into two chambers by a 10-kDa regenerated cellulose membrane. One side of each dialysis cell was dosed with 7.97 mg HSA ( $\leq 0.02\%$  fatty acids, Sigma-Aldrich) in phosphate-buffered saline (PBS, pH 7.4 prepared in HPLC grade water) to a final concentration of 600  $\mu\text{M}$ , which mimics physiological conditions (Myatt, 2017). The other side of each cell was then dosed with 200  $\mu\text{L}$  of an AFFF dilution (4000 to 16,000-fold in PBS) or an in-house mixture of 26 PFAS (prepared with PFAC-24PAR, PFPrS, and br-PFOS from Wellington Laboratories Inc.). Seventeen out of the 26 target PFAS in this study are commonly measured in drinking water using EPA method 533 or EPA method 537.1.

An aliquot of legacy AFFF (3M, 1999) was provided by Professor Christopher Higgins at the Colorado School of Mines. Experimental batches consisted of six concentration levels of AFFF or PFAS standard dilutions and were replicated four times. For each batch of experiments, a method blank was prepared with HSA free of AFFF or PFAS. A negative control cell containing AFFF or the PFAS standard mix was prepared without HSA to assess free movement of PFAS

through the membrane (Figure B1.1). The system was incubated at 37 °C while rotating at 30 RPM for 108 h (Figure B1.2).

### *3.3.3 PFAS Extractions*

Aliquots from each dialysis cell were processed to generate three extracts, shown in Figure 1. These were (1) free PFAS from the aqueous fraction, (2) noncovalently bound PFAS associated with the dissolved protein, and (3) residual PFAS in the precipitated protein pellet (candidates for ultrastrong noncovalently bound or potentially covalently bound PFAS). Briefly, post-dialysis aqueous samples (100 µL) from the chemical chambers were equilibrated with 50% methanol. Protein aliquots (100 µL) of the post-dialysis protein cells were extracted with formic acid (FA) acidified acetonitrile (ACN) for protein denaturation and precipitation. Noncovalently bound PFAS were determined by taking the concentration difference between protein aliquot fraction and chemical fraction. The residual protein pellets were further washed with 1 mL ACN five times, and the last wash was concentrated (to 200 µL) and saved to verify the absence of PFAS. We then applied a standard acid hydrolysis protocol for amino acid analysis to break peptide bonds and to release any PFAS that were possibly covalently bound to HSA (Hirs et al., 1954; Lapierre et al., 2019; Muñoz et al., 2011; Mustăţea et al., 2019; Otter, 2012). Extracts of noncovalently bound PFAS and residual PFAS in the protein pellet were solvent exchanged into 50% methanol to match the final solvent composition of the aqueous extracts. Additional details on the sample preparation can be found in Appendix B (B1.1). An internal standard mix (ISTD, Table B6.1) in 50% methanol was added into each extract prior to analysis. Details on instrumentation and acquisition settings are provided in Table B5.1.

### *3.3.4 Analytical Instrumental Set-up*

Quantification of 26 PFAS targets and qualification of diverse PFAS via suspect screening were performed on each AFFF dilution and each PFAS extract from dialysis. Data were acquired on an Agilent 1260 Infinity HPLC system paired with a 6530 QTOF MS. Sample extracts (10  $\mu$ L) were injected onto a C18 column (ZORBAX RRHD Eclipse Plus C18 column; 2.1 mm  $\times$  150 mm, 1.8  $\mu$ m, Agilent Technologies, Inc.) at a flow rate of 0.40 mL/min, with a total run time of 31.5 min. The aqueous mobile phase (A) was 20 mM ammonium acetate (Fisher Scientific) in Optima™ HPLC grade water (Fisher Scientific) and the organic mobile phase (B) was 100% Optima™ HPLC grade acetonitrile (Fisher Scientific). The mass spectrometer ionized samples in a negative mode using collision energies (CE) of 0, 10, 20, and 40 eV. A quality-control run of the 26-PFAS standard mix was analyzed after every 8 samples to ensure that concentrations of targeted PFAS remained within 30% of known concentrations.

### *3.3.5 Suspect Screening*

For suspect screening, mass-to-charge-ratios ( $m/z$ ) of 50–1200 were fragmented in the collision cell with CE of 0, 10, 20, and 40 eV in All-Ions acquisition mode. Data were processed using Agilent MassHunter Qualitative Analysis (B.08.00) by applying the “Find by Formula” search against an in-house AFFF Personal Compound Data Library (PCDL). The PCDL contained 3,793 PFAS extracted from the Norman Suspect List Exchange (OECDPFAS, 2018b), and 727 hydrocarbon surfactants (monoisotopic mass: 100–1200) extracted from the surfactant suspect list curated by Schymanski et al. (2014). The PCDL also included 63 MS/MS spectra, of which 31 spectra were acquired from in-house standards, 24 spectra were extracted from MassBank (Horai et al., 2010), and six spectra were generated with CFM-ID 3.0 (Allen et al., 2014). PFAS were considered qualified with level 2–3 confidence as outlined in Schymanski et al. (2014). Suspect-screening search settings are listed in Table B5.2. Qualified PFAS were reported only if the

following additional criteria were met: (1) the abundance of the qualified ion was greater than three times the experimental blank ion abundance; (2) the ion was qualified in at least 67% of dialysis cells in each batch of experiments; and (3) the PFAS qualified in dialysis cell extracts were also qualified in neat AFFF dilutions. All qualified PFAS that met these additional criteria were confirmed by re-running extracts with data-dependent targeted analysis (see Section 2.6).

### 3.3.6 Targeted MS/MS

For data-dependent targeted analysis, a list of targeted mass-to-charge ratios (m/z) and corresponding retention times (RT) was compiled based on pre-runs with All-Ions acquisition described in Section 2.5. To avoid overlapping peaks of targeted compounds, duplicate injections were performed for each sample to ensure at least 0.4 min RT difference between peaks of any two targeted compounds within one injection. This approach can minimize false identifications of PFAS due to the instrument's inherent mass error.

### 3.3.7 PFAS Quantification

Quantification was performed with 19-ISTD dosed ten-point calibration curve (0.5–250 ng/mL). Whole method limits of quantification (LOQ) range from 0.025 to 1 ng/mL. Details of analytical standards and extraction recoveries are available in Table B6.1

### 3.3.8 Experimental Determination of PFAS Noncovalent Binding Affinities

The concentrations of 26 PFAS targets that partitioned into the protein chamber, remained in the chemical chamber, and remained associated with the protein pellet were directly determined by HPLC-QTOF-MS. Noncovalent binding affinities, measured as association constants ( $K_A$ ), were calculated assuming a one site specific binding as shown in Equations (1) and (2).



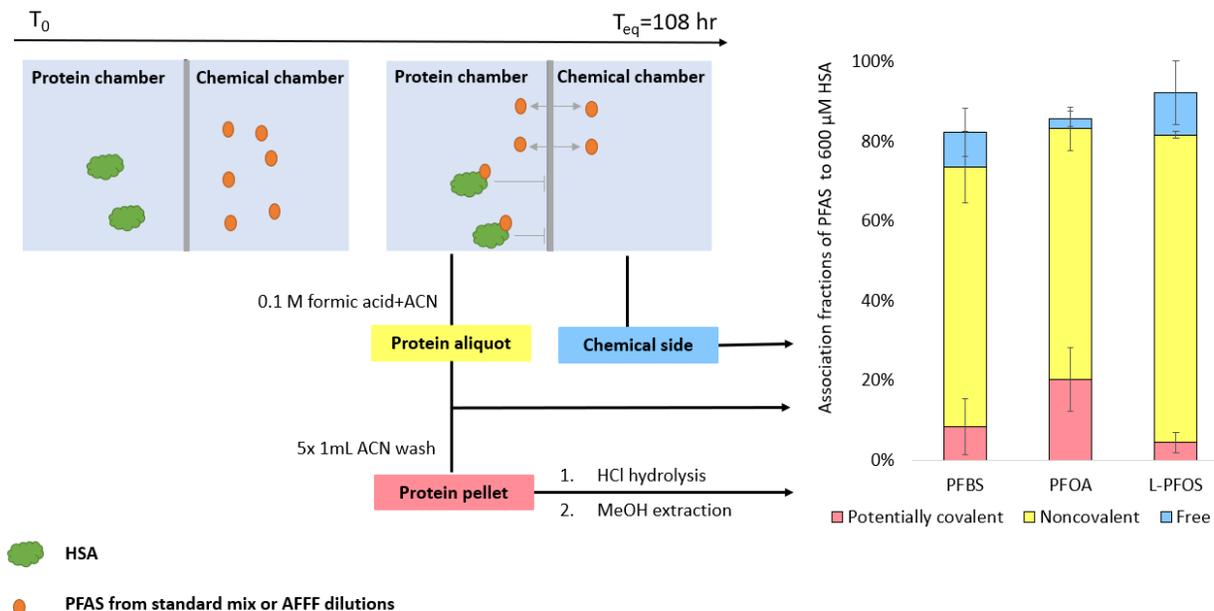
$$\frac{[\text{HSA-PFAS}_i]}{[\text{HSA}][\text{PFAS}_i]} = \frac{k_{on,i}}{k_{off,i}} = K_{A,i} \quad (2)$$

To enable the assessment of multiple specific binding sites, we also fitted the data to the Langmuir isotherm model (X. Zhang et al., 2009) with a limited binding sites assumption, following Equations (3) and (4).

$$q_m = [\text{HSA-PFAS}_i] + q_{0,i} \quad (3)$$

$$\frac{1}{[\text{HSA-PFAS}_i]} = \frac{[\text{PFAS}_i] * q_{0,i}}{\frac{1}{K_{A,i}} + [\text{PFAS}_i]} \quad (4)$$

In these equations,  $i$  refers to the compound of interest,  $q_m$  is the concentration of total binding sites, and  $q_{0,i}$  is the concentration of empty binding sites.  $[\text{PFAS}_i]$  is the concentration of free PFAS<sub>*i*</sub> measured in the chemical side of the equilibrium dialysis set-up.  $[\text{HSA-PFAS}_i]$  is calculated by taking the difference between the concentration of PFAS in the protein side and in the chemical side, as shown in Figure 6. Additional isotherm models, including linear adsorption and Freundlich adsorption models, were also evaluated (Figure B1.3).



**Figure 6.** Equilibrium dialysis set-up and observed mass balance for three PFAS with initial dosages of 40–80 ng. PFAS that were free in aqueous solution PFAS, noncovalently bound PFAS, and residual PFAS in the precipitated protein pellet were measured independently. The time required to reach equilibrium ( $T_{eq}$ ) was previously determined using 26 PFAS standards in Figure B1.2. Detailed mass balance data are available in Table B1.1.

### 3.3.9 Computational Simulations of Noncovalent PFAS Protein Binding

We used AutoDock Vina (v 1.1.2) (Trott & Olson, 2010) to dock 62 ligands (26 PFAS targets, 18 qualified PFAS, and 18 qualified hydrocarbon surfactants in AFFF) to two HSA crystal structures (Protein Data Bank entries 1E7G and 1AO6). 1E7G was chosen as the native structure of HSA, which complexed with tetradecanoic acid (myristic acid) (Bhattacharya et al., 2000; Wardell et al., 2002). 1AO6 is an unliganded HSA structure and may be more similar in conformation to the HSA we used experimentally since the protein standard we purchased contained low levels of fatty acids (<0.02%). We followed the workflow outlined by Ng and Hungerbuehler (2015) with several modifications. Specifically, in the ligand-preparation step, we used the “Generating Conformers” function in DataWarrior V5.2.1 (Sander et al., 2015) to generate 3D structures for all ligands. Then, we optimized ligand structures using the MMFF94s forcefield in Avogadro V1.90.0 (Hanwell et al., 2012). In addition, we used PyMol (v2.3.3) for structure visualization, redocking alignment, and crucial residue identification (Schrodinger LLC, 2015; Seeliger & de Groot, 2010). Simulations were repeated 100 times for 6 binding pockets, and each simulation generated 9 binding modes, yielding 5400 predictions in total for each PFAS. Further details on the docking method as well as simulation precision and accuracy are available in Appendix B (B3.1 and B3.2).

The simulation method was evaluated by redocking PFOS on an experimentally determined HSA (Protein Data Bank entry 4E99) structure that was originally complexed with two PFOS in fatty-acid binding site (FA) 3/4 and 5. The atomic root-mean-square-deviation (RMSD) of redocked PFOS on FA 3/4 and FA 5 was determined to be less than two angstroms, indicating successful redocking. The redocking search information and RMSD statistics are available in Table B2.3.

## 3.4 Results and discussion

### 3.4.1 Characterization of PFAS in AFFF

Targeted analysis using 26 PFAS standards was insufficient for characterizing the AFFF sample: less than 9% of the total organic fluorine was quantified as compared to quantitative  $^{19}\text{F}$  NMR (Table B7.1). In addition to the target PFAS, we identified 18 other PFAS and 18 hydrocarbon surfactant structures using suspect screening analysis for initial qualification. The suspected PFAS and hydrocarbon surfactant structures were further confirmed via data-dependent acquisition or library spectrum match. Manual annotation of the MS/MS spectra supported identification of these compounds (B1.2, Figure B4.1 to B4.19). Based on structural categorizations conducted by the OECD (2018a), PFAS qualified in the AFFF sample included: 19 perfluoroalkane sulfonyl compounds, seven perfluoroalkyl carbonyl compounds, four fluorotelomer-related compounds, and two side-chain fluorinated aromatic compounds. Eight PFAS suspects were qualified in the initial screening but eliminated via manual confirmation. A full list of the hydrocarbon surfactants identified in AFFF is provided in Table B6.2. These hydrocarbon surfactants homologous series were detected using EnviHomolog (<http://www.envihomolog.eawag.ch>). A repeating mass-increment of 14.0156 ( $-\text{CH}_2-$ ) was observed for four Linear Alkylbenzyl Sulfonates (LAS). Repeating mass-increments of 28.0313 ( $-\text{C}_2\text{H}_4-$ ) and 44.0262 ( $-\text{C}_2\text{H}_4\text{O}-$ ) were observed for 31 Alkyl Ethoxy Sulfates (AES).

Most PFAS identified in this study were also reported by other studies (Houtz et al., 2013; Barzen-Hanson et al., 2017; and McDonough et al., 2020) for the same or similar AFFF commercial products. We qualified  $\text{C}_4$  and  $\text{C}_6$  perfluoroalkyl sulfonamide amino carboxylates and perfluoroalkyl sulfonamido amines that were reported by Houtz et al. One pentafluorosulfide-containing eight perfluorocarbon PFAS (8-F5S-PFOS) found in AFFF was also reported by

Barzen-Hanson et al. We qualified four of six PFOS-substituted compounds (H-PFOS, U-PFOS, Cl-PFOS, and K-PFOS) and one of two PFDS-substituted compounds (H-PFDS) reported by McDonough et al. A C<sub>6</sub> containing phosphonic acid and ester functional groups (8:2 monoPAP-diEes) identified in this study was detected in PFAS-contaminated soil (from paper sludge) in Germany (Bugsel & Zwiener, 2020). We identified six novel PFAS that have not been otherwise detected in environmental samples to our knowledge: 4-FHp-CycHxA, diTF-IsoBA, Hx-diFB, Uridine-FB, and two C<sub>4</sub> fluoroalkyl esters (N-PFBS-MFPe, N-FBEAc). Fluoroalkyl esters may undergo hydrolysis in the ambient environment and eventually release PFSA or PFCA (C. J. Norwegian E. A. Nielsen, 2017).

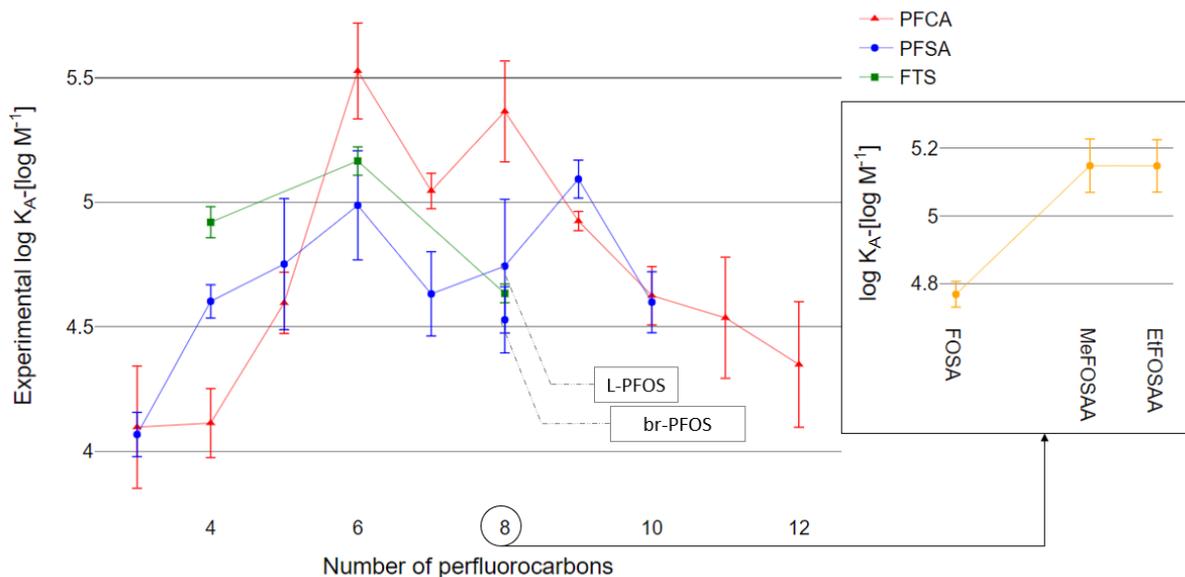
### *3.4.2 Noncovalent Binding of PFAS in AFFF to Human Serum Albumin*

Of 32 PFAS identified in the AFFF, 28 PFAS bound noncovalently to HSA in equilibrium dialysis experiments. We confirmed 14 of these PFAS with analytical standards, and the remaining via data-dependent acquisition. Five PFAS were qualified and confirmed in the precipitated and washed protein pellets. Since PFAS released from hydrolyzed HSA pellets could not be extracted with the organic solvent, natural dissociation of this fraction of PFAS was not expected in a reasonable timeframe. Hence, these PFAS were considered as candidates for ultrastrong noncovalent or potentially covalent binding to HSA. Fourteen additional PFAS were qualified (confidence level 4) in the protein pellet, but were not qualified in the AFFF dilutions. We excluded these compounds from further analysis.

Consistent with previous studies of PFAS-protein associations, PFAS were highly bound to HSA (Allendorf et al., 2019; H. N. Bischel et al., 2011a; X. Zhang et al., 2009). HSA contains multiple PFAS binding sites with potentially different binding affinities (H. N. Bischel et al., 2010; Chi et al., 2018) such that measured K<sub>A</sub> values represent a mixture of affinities for different binding

sites. A majority of the PFAS exhibited linear binding isotherms, indicating nonspecific noncovalent associations with HSA (Figure B1.4 and Table B1.1).  $K_A$  followed an inverted-V trend by which  $K_A$  increased with perfluorocarbon chain elongation up to C<sub>6</sub> through C<sub>9</sub> and subsequently decreased (Figure 7). The trend for C<sub>4</sub> through C<sub>6</sub> and C<sub>8</sub> through C<sub>11</sub> PFCAs is consistent with the pattern for bovine serum albumin (BSA)–water distribution coefficients ( $K_{PW}$ ) previously determined (H. N. Bischel et al., 2011a). The PFSA trend for C<sub>4</sub> through C<sub>8</sub> is consistent with the BSA-association constants determined by Allendorf et al. (2019). Our measurements of  $K_A$  were generally an order of magnitude lower than  $K_A$  measured by Allendorf et al. (2019), with the exception of PFBA, PFHpA, and C<sub>6</sub> PFAS. We have greater confidence in the physiological relevance of our experimental results as our results were obtained at physiologically relevant molar ratios of PFAS and HSA, and our  $K_A$  values were determined from isotherm data rather than single-point experimentation. As low levels of AFFF exposures to humans are most common (Kaboré et al., 2018), we tested PFAS-HSA association constants from 0.001 to 0.1 PFAS: HSA.

Overall, a trend of increasing  $K_A$  with perfluorocarbon-chain length was observed for PFCAs and PFSAs up to C<sub>6</sub>. HSA binding affinities of perfluorohexanesulfonic acid (PFHxS) and perfluoroheptanoic acid (PFHpA) were exceptionally high (Log  $K_A$ :  $4.99 \pm 0.44$  and  $5.53 \pm 0.39$ , respectively). These observations are consistent with long blood plasma elimination half-lives reported for PFHxS in humans (Y. Li et al., 2018; Olsen et al., 2007) and for PFHpA and PFHxS in pigs (F. Yang et al., 2014).



**Figure 7.** Experimentally determined noncovalent association constants ( $K_A$ ) of 26 PFAS targets with human serum albumin (has) in equilibrium dialysis. Three C<sub>8</sub> precursor compounds, Perfluoro-1-octanesulfonamide (FOSA), N-methylperfluoro-1-octanesulfonamidoacetic acid (MeFOSAA), and N-ethylperfluoro-1-octanesulfonamidoacetic acid (EtFOSAA) are presented in a separate plot. The error bars represent one standard deviation.

### 3.4.3 Residual PFAS in Precipitated Protein Pellet Covalently Binding to HSA

In HSA binding experiments using the 26 PFAS targets, three PFAS (PFBS, PFOA, and PFOS) were consistently quantified in the protein pellets. The protein pellets contained 7% PFBS, 20% PFOA, and 5% PFOS of the total spiked mass (80 ng) of each of these compounds (Figure 6). PFHxS was detected inconsistently in the protein pellets (11 out of 24 samples among four trials). The AFFF (4E3 diluted with PBS buffer)-spiked protein pellets contained 1% of spiked PFBS, 26% of spiked PFOA, and 2% of spiked PFOS. N-(3-(dimethylamino)propyl)-1,1,2,2,3,3,4,4,5,5,6,6,6-tridecafluorohexane-1-sulfonamide FHxSA (N-diMAmP-FHxSA) and 4,4,4-trifluoro-2-(2,2,2-trifluoroethyl)butanoic acid (diTF-IsoBA) were also detected consistently in the protein pellets (at least 16 out of 18 pellets among 3 trials) after AFFF exposure, but could not be quantified due to lack of available standards.

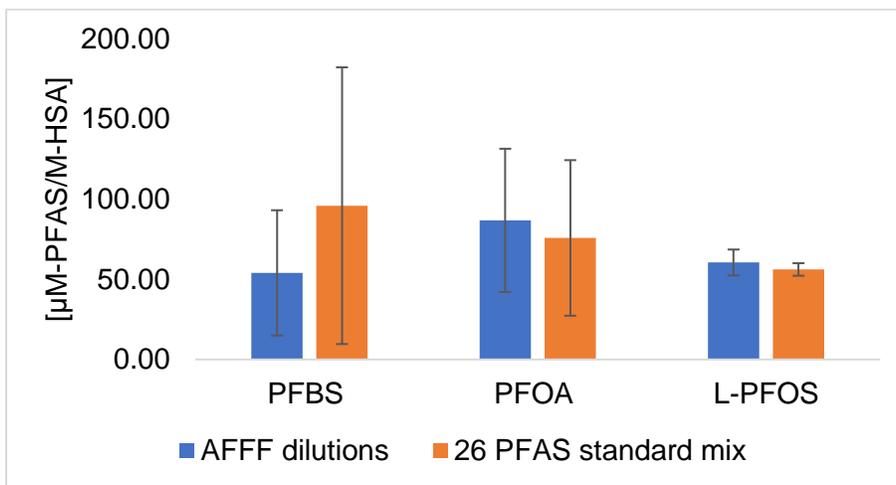
Consistent detection of three PFAS in the digested protein pellets precipitated from AFFF and PFAS standard exposure experiments is presented in Figure 4. Residual protein pellets from HSA exposed to PFAS standards (which did not include C<sub>4</sub> precursors) contained similar amounts of PFBS, PFOA, and PFOS as protein pellets from HSA exposed to AFFF dilutions (Figure 8 and Table B1.1). PFAS release from the protein pellet could be a result of several factors. First, residual PFAS in the pellet could be present as an analytical artifact resulting from high-concentrations of PFAS in AFFF. However, no quantifiable level of PFAS was observed in the protein pellet washes, blank cells, or control cells used in the equilibrium dialysis experiment (See Figure B1.1). Additionally, AFFF was diluted from 2000-fold to 80,000-fold prior to HSA exposure in the most dilute case, and the PFAS targets were still present in the precipitated protein from these tests. Second, PFAS could be retained in the pellet via ultrastrong noncovalently interactions, which was observed in an amphiphilic polymerization system (Krieg et al., 2014). In conjunction with an

exterior aqueous environment, large proteins like HSA that have multiple hydrophobic binding sites can provide an amphiphilic environment in which different protein residues interact with the polar headgroups and hydrophobic perfluorotails of PFAS simultaneously (F. Yang et al., 2014). Third, PFAS could be retained in the pellet via covalent interactions, the potential for which we evaluate in further detail below.

To our knowledge, only two types of PFAS have been reported to bind covalently to proteins: PFCAs (vanden Heuvel et al., 1992) and fluorotelomer unsaturated aldehydes (FTUALs) (Rand & Mabury, 2012, 2013). The mechanism of covalent binding between PFAS and proteins remains unclear, but thiol- and nitrogen-containing nucleophilic amino acids in serum albumin proteins are suspected to play a role (vanden Heuvel et al., 1992). In the case of FTUALs, covalent bond formation occurs via Michael addition (Rand & Mabury, 2013). This mechanism cannot explain our observation of perfluoro alkyl carboxylic and sulfonic acids in the protein pellet. Formation of covalent bonds between carboxylic acids or sulfonic acids containing ligands and protein residues has not been observed under physiological conditions (Dietzen, 2018; Wilbur, 2003). The low mole ratio of residual PFAS detected in the protein pellet to initial HSA levels indicates that not all PFAS-HSA associations resulted in PFAS retention in the protein pellets.

While we are unable to disentangle the mechanisms explaining PFAS in the protein pellet, we consider residual PFAS in protein pellets as candidates for ultrastrong noncovalent or potentially covalent binding to HSA. In addition to the perfluoro alkyl acids (PFAAs) described above, we noted that 14 additional PFAS qualified in the dialysis extracts were not further evaluated in this study (i.e., by acquiring targeted MS/MS). It is possible that these PFAS were generated from reactions with HSA or through transformations during the acid-hydrolysis processing step. We would expect the formation of PFAS-HSA covalent bonds and subsequent

digestion of PFAS-containing HSA to yield PFAS-peptide complexes. Future analysis should evaluate whether perfluorocarbon moieties are associated with amino acids or peptides following digestion.



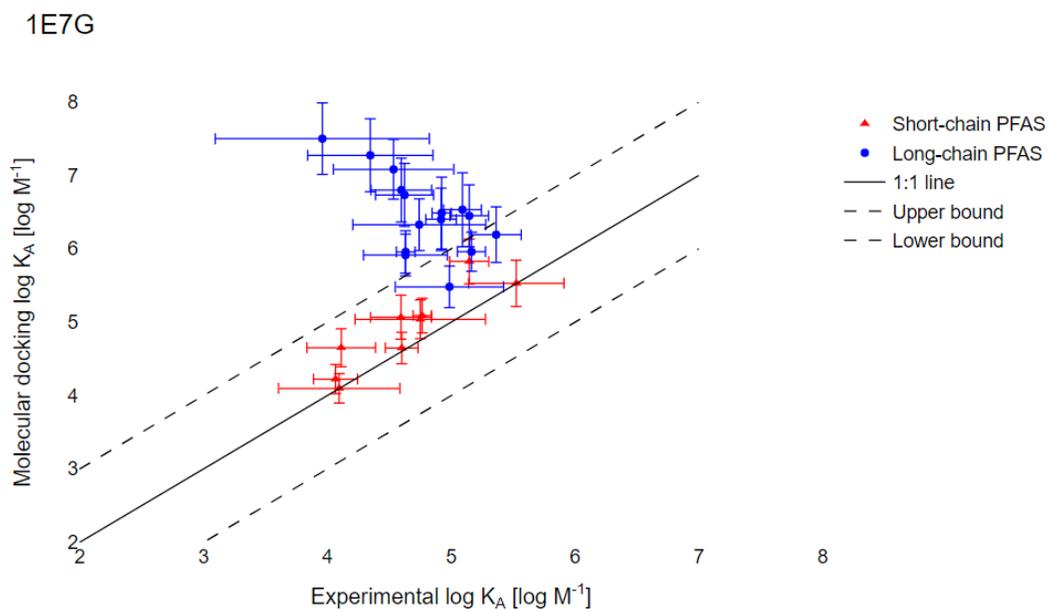
**Figure 8.** Residual PFAS quantified in digested HSA pellets, precipitated from equilibrium dialysis experiment.

#### 3.4.4 Quantitative Determination of PFAS–HSA Association Constants

We experimentally quantified HSA binding affinities for 26 PFAS: 11 PFCAs (C<sub>3</sub> through C<sub>13</sub>), nine PFSA (C<sub>3</sub> through C<sub>12</sub>), one perfluoroalkane sulfonamide (C<sub>8</sub>), two perfluoroalkane sulfonamide acetic acids (methylated and ethylated C<sub>8</sub>), and three fluorotelomer sulfonic acids (4:2 FTS, 6:2 FTS, and 8:2 FTS). PFAS-HSA association constants ranged from 10<sup>4.0</sup> to 10<sup>5.5</sup> M<sup>-1</sup> (Figure 7).

#### 3.4.5 Evaluation of Molecular Docking to Predict the PFAS–HSA Binding Affinities

Molecular docking of PFAS with two HSA crystal structures (1E7G and 1AO6) was used to simulate K<sub>A</sub> for the 26 PFAS tested experimentally (Table B7.2). Accurate K<sub>A</sub> predictions using 1E7G were limited to short-chain PFAS. In Figure 9, for the nine short-chain PFAS (PFCA with less than six perfluorinated carbons and PFSA with less than five perfluorinated carbons), a significant positive correlation between the docking-predicted K<sub>A</sub> and the experimentally determined values was observed (95% CI: slope = 1.02 ± 0.19, r = 0.900). For the 17 long-chain PFAS, significant negative correlation between the docking-predicted K<sub>A</sub> and the experimentally determined values was observed (95% CI: slope = -1.05 ± 0.30, r = 0.7680).



**Figure 9.** Comparison of experimental  $\log K_A$  with results from molecular simulations with the HSA crystal structure 1E7G. Solid black lines represent the 1:1 line; dotted lines represent one log unit higher or lower. Error bars reflect one geometric standard deviation (GSD).

### 3.4.6 Semi-quantification of Bioconcentration Factors of Qualified PFAS

We calculated pseudo-bioconcentration factors ( $BCF_{\text{pseudos}}$ ) for noncovalently bound and potentially covalently bound fractions separately to evaluate patterns related to perfluorocarbon chain length and functional groups (Figure 10).  $BCF_{\text{pseudo}}$  serves as a quantitative benchmarking technique to cross-compare bioaccumulation potentials of novel PFAS using qualitative screening data (McDonough et al., 2020). The  $BCF_{\text{pseudo}}$  was calculated for each PFAS as follows:

$$BCF_{i,pseudo} = \frac{A_{i,sample}}{A_{L-PFOS,sample}} * \frac{A_{L-PFOS,AFFF}}{A_{i,AFFF}} \quad (5)$$

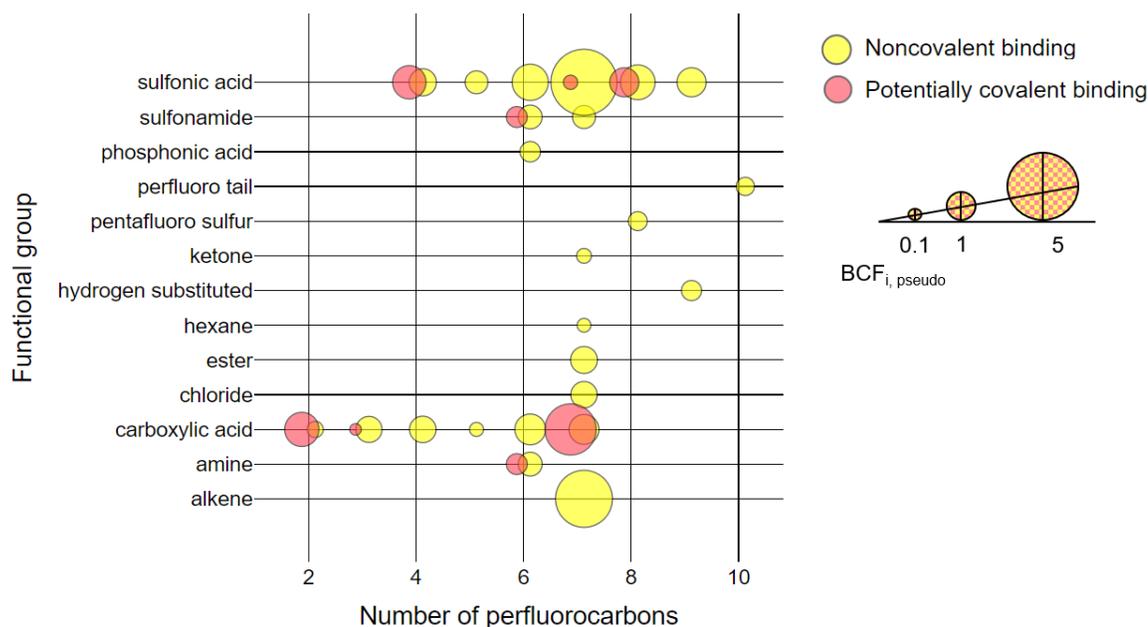
where  $A_{i,sample}$  is the peak area of compound  $i$  detected in the protein aliquot or pellet;  $A_{L-PFOS,sample}$  is the peak area of L-PFOS detected in the protein aliquot or pellet;  $A_{L-PFOS,AFFF}$  is the peak area of L-PFOS detected in neat AFFF dilutions; and  $A_{i,AFFF}$  is the peak area of compound  $i$  detected in neat AFFF dilutions. All peak areas were normalized with their corresponding internal standard peak area prior to the calculations.

Analysis of  $BCF_{\text{pseudo}}$  revealed several key findings in Figure 10. First, sulfonic acids ( $C_4$  through  $C_9$ ) and carboxylic acids ( $C_2$  through  $C_7$ ) consistently exhibited high  $BCF_{\text{pseudos}}$  in the noncovalently bound fraction. Second,  $C_7$  PFAS across different functional groups consistently exhibited high  $BCF_{\text{pseudos}}$  in the noncovalently bound fraction. We observed a noncovalent binding trend with a turning point at  $C_7$  for all qualified PFAS, which is consistent with observations for targeted PFASs (Figure 7). Third, PFOA exhibited the highest  $BCF_{\text{pseudo}}$  for potentially covalently bound fractions. This is despite low levels in the AFFF; PFOA represented less than two percent of the total organic fluorine mass in the AFFF among all PFAS compounds quantified through targeted analysis. Finally, three PFAS with four or fewer perfluorocarbons exhibited higher  $BCF_{\text{pseudos}}$  in the potentially covalently bound fraction than L-PFOS. The  $BCF_{\text{pseudo}}$  for the  $C_2$  carboxylic acid in the potentially covalently bound fraction ( $BCF_{\text{diTF-IsoBA, pseudo}} = 13.8 \pm 3.14$ ) was

an order of magnitude greater than the  $BCF_{\text{pseudo}}$  for L-PFOS. This finding is concerning, as short-chain PFAS are typically considered less bioaccumulative than long-chain PFAS and exhibit weaker noncovalent interactions with proteins (Danish Ministry of the Environment, 2015). Our results indicate that short-chain PFAS may in fact be strongly retained in proteins (and in precipitated protein pellets) even when present at low concentrations in serum. Further studies should consider the impacts of these observations on analytical conclusions as well as potential toxicological risks.

To assess the ability of in-vitro binding studies with HSA to represent bioaccumulation potentials of PFAS in animals, we compared our  $BCF_{\text{pseudos}}$  from the noncovalently bound fraction to pseudo-bioaccumulation factors ( $BAF_{\text{pseudos}}$ ) calculated in an in-vivo mouse-dosing study that used the same AFFF commercial product (McDonough et al., 2020). Our calculation of  $BCF_{\text{pseudo}}$  (Equation 5) was equivalent to the calculation of  $BAF_{\text{pseudo}}$  by McDonough et al. (2020). However, we performed direct exposure of HSA to PFAS while McDonough et al. used a PFAS sample from mouse serum following AFFF oral gavage. Twelve types of PFAS (43 distinct chemical structures) were qualified by McDonough et al. in the mouse serum (which excluded analysis of the protein pellet and associated ligands) following oral gavage of the AFFF. Of the six reported noncovalent  $BAF_{\text{pseudos}}$  for PFOS substitutes prevalent in mouse serum, two were similar to the noncovalent  $BCF_{\text{pseudos}}$  we observed noncovalently bound to HSA. The bioconcentration potential of U-PFOS with seven perfluorocarbons was high in both studies ( $BCF_{\text{U-PFOS,pseudo}} = 8.85 \pm 0.16$  with HSA, compared to  $BAF_{\text{U-PFOS,pseudo}} = 6.7$  in mouse serum, averaged across genders). The bioconcentration potential of noncovalently associated Cl-PFOS ( $BCF_{\text{Cl-PFOS,pseudo}} = 0.82 \pm 0.16$  for HSA) was similar to L-PFOS ( $BCF_{\text{L-PFOS,pseudo}} = 1$ ) in both McDonough et al. and this study. Comparisons between the two studies may otherwise aid in

identifying products of metabolic biotransformation. For example, the mouse serum  $BAF_{\text{pseudo}}$  for hydrogen-substituted PFAS (H-PFOS) in McDonough et al. was about one order of magnitude higher than the  $BCF_{\text{H-PFOS,pseudo}}$  we observed for HSA. We suspect that biotransformation—which only takes place for in-vivo experiments—might have contributed to the high  $BAF_{\text{H-PFOS,pseudo}}$  observed by McDonough et al., given that H-PFOS is a daughter product of precursor PFAS in AFFF. Altogether, these results indicated the value of comparing  $BAF_{\text{pseudo}}$  and  $BCF_{\text{pseudo}}$  to assess contributions of protein binding and metabolism of PFAS precursors as explanatory factors for PFAS bioaccumulation.

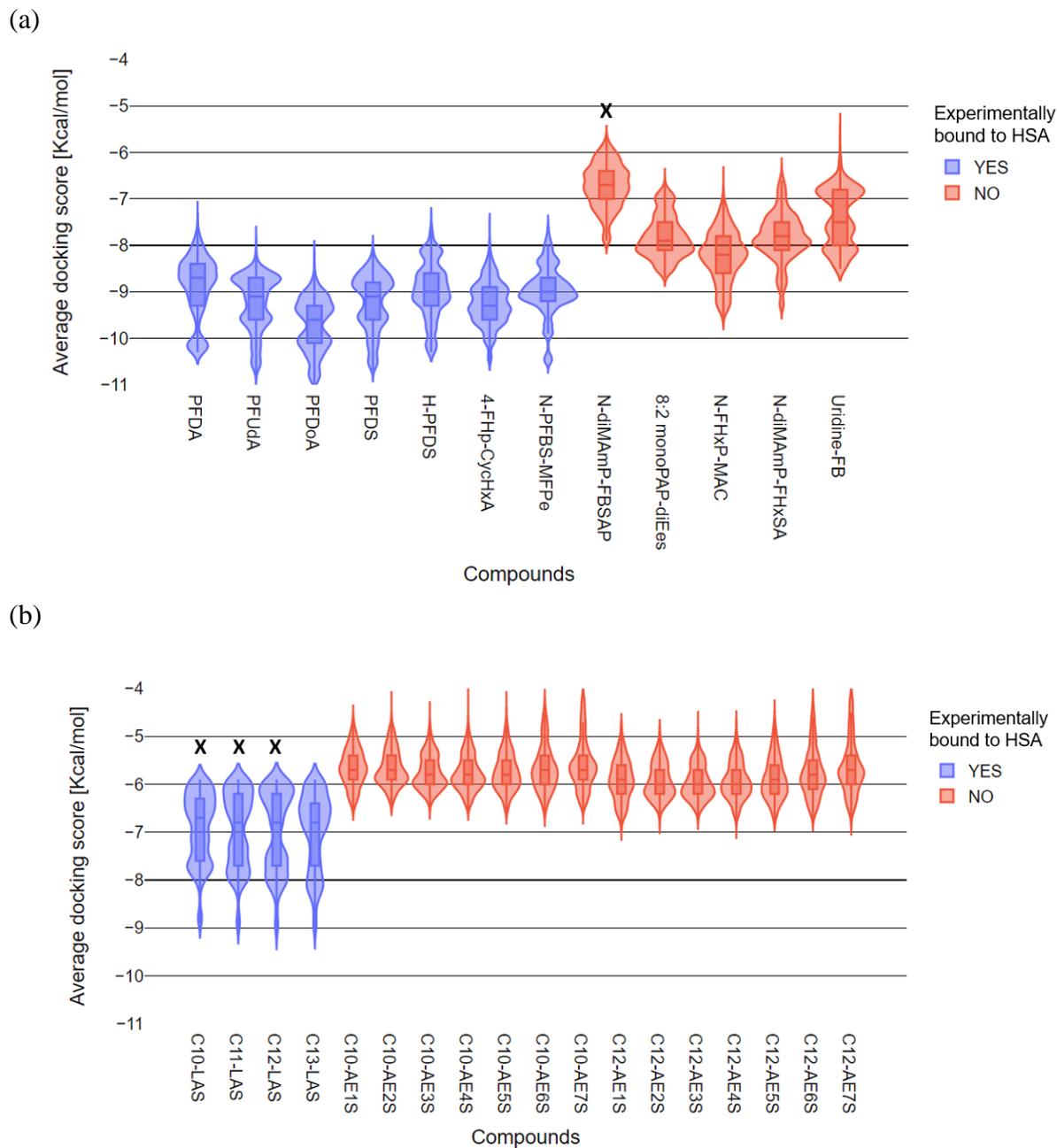


**Figure 10.** Pseudo-bioconcentration factors ( $BCF_{i,pseudo}$ ) of PFAS in aqueous film-forming foams (AFFF) that were bound noncovalently (yellow) or potentially covalently (red) to HSA. Bubble size represents the natural logarithm of the  $BCF_{pseudo}$  (legend  $BCF_{pseudo} = 1$ ). For 12 qualified PFAS with multiple functional groups, a separate bubble of the same size is displayed for each functional group (e.g., noncovalently bound Cl-PFOS contained chloride and sulfonic acid groups and is represented as two bubbles with seven perfluorocarbons).

### 3.4.7 Qualitative Prediction of HSA-bound vs. Nonbound Compounds

While docking scores are unreliable for quantitative predictions of binding affinities, docking scores can be used for qualitative comparisons between complexes and to identify candidate ligands for the protein of interest. We used AutoDock Vina to identify HSA-binding compounds in AFFF by comparing PFAS and hydrocarbon surfactants in AFFF. Results of docking are presented as violin plots, which display the distribution of simulated docking scores (Figure B3.1 to B3.6). For each PFAS, a kernel density plot of docking scores was derived using the 5400 conformations generated from simulations on six HSA binding pockets. All hydrocarbon surfactants identified in AFFF had more than 10 carbons in the backbone, so we selected docking results for PFAS with 10 or more perfluorocarbons in the backbone for comparison (Figure 11). For both sets of long-chain PFAS and hydrocarbon surfactants, we observed a separation of distributions that appeared to distinguish binding (low docking scores) or non-binding (high docking scores) compounds. The low and high docking scores corresponded, respectively, to PFAS that were observed as bound or unbound to HSA in in-vitro experiments. We performed a Kruskal–Wallis test of docking predicted binding scores for the compounds in Figure 6. Most compounds in AFFF that we observed to bind to HSA in the in-vitro experiments also had predicted docking scores that were significantly different from the unbound compounds ( $p < 0.05$  in Table B7.3). However, the binding score of one experimentally determined unbound PFAS (N-diMAmP-PBSAP) and three experimentally determined bound hydrocarbon surfactants (C10-LAS, C11-LAS, and C12-LAS) were not significantly different from each other ( $p = 1$ , in Table B7.3, labeled with black X in Figure 11). Experimental results for suspect screening were thus largely consistent with simulated docking scores when comparing HSA binding affinities within the same class of chemicals. The shape of the kernel density plots may also provide insights into

different binding processes. The kernel density plots for both bound and unbound PFAS (Figure 6a, B2.5, B2.6, and B2.7) are more varied in shape than those for hydrocarbon surfactants (Figure 5b, B3.0). Similar to kernel plots, cluster analyses were commonly used to identify preferential ligand binding sites, suggesting site-specific binding between PFAS and HSA (Hamdi et al., 2015; Salvalaglio et al., 2010). Predicted HSA binding scores for bound hydrocarbons exhibited bimodal or even trimodal (C10-LAS and C11-LAS) distributions. Predicted HSA binding scores for nonbound hydrocarbons converged and centralized for all HSA binding pockets (Figures B2.8 and B2.9). To further validate and accurately predict PFAS-HSA binding energies, mechanistic studies using molecular dynamics coupled with molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) methods for these PFAS are ongoing in our research group.



**Figure 11.** Violin plots of molecular docking simulated 1E7G docking scores for (a) PFAS with greater than 10 perfluorocarbons and (b) hydrocarbon surfactants identified in AFFF that have greater than 10 carbons. The shape of each violin represents a rotated kernel density plot of 5400 HSA–PFAS binding conformations generated from simulations for six binding pockets. Blue and red colors are used to distinguish experimental results. The compounds with significantly greater peak area (after correction with ISTD peak area) in the noncovalently bound fraction of the protein chamber relative to the chemical chamber are shown in blue. The compounds identified experimentally in the protein chamber that were not significantly greater in peak area relative to the chemical chamber are shown were red. Distributions in red were significantly different than distributions in blue Kruskal–Wallis ( $p < 0.05$ ) except for those distributions marked with a black X.

### 3.5 Conclusion

This study explored the value of suspect screening and computational simulations to identify potentially bioaccumulative PFAS from a PFAS-containing commercial product, AFFF. A majority of the PFAS we identified in a legacy AFFF bind to the most abundant protein in human serum, human serum albumin (HSA). At least five PFAS, including two PFAS with less than five perfluorocarbons, were detected in the precipitated and washed protein pellet. The potential health implications of ultrastrong or covalent binding of PFAS are unclear. Covalent modifications of HSA affect the clearance and metabolic destiny of many drugs, and have been hypothesized as the center of toxicity exhibited by many drugs.

Our observation of binding of short-chain PFAS to albumin is concerning and requires further mechanistic assessment. Short-chain PFAS are largely considered less bioaccumulative, with shorter half-lives in organisms, than long-chain PFAS. Our results indicated that some short-chain PFAS may be retained in the blood for much longer—these PFAS remained associated with HSA after extensive solvent washing. Computational simulations for bioaccumulation potential can provide value by decreasing reliance on time- and labor-intensive laboratory experiments. Though predicted binding scores cannot quantitatively describe binding affinities with HSA, the scores can be used to qualitatively identify previously uncharacterized PFAS that are likely to bind to HSA. More broadly, this study offers a framework for evaluating bioaccumulation potentials of thousands of PFAS in comparable biological tissues.

### 3.6 References

- Allen, F., Pon, A., Wilson, M., Greiner, R., & Wishart, D. (2014). *CFM-ID : a web server for annotation , spectrum prediction and metabolite identification from tandem mass spectra*. *12*(June), 94–99. <https://doi.org/10.1093/nar/gku436>
- Allendorf, F., Berger, U., Goss, K. U., & Ulrich, N. (2019). Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. *Environmental Science: Processes and Impacts*, *21*(11), 1852–1863. <https://doi.org/10.1039/c9em00290a>
- ATSDR. (2017). *An Overview of Perfluoroalkyl and Polyfluoroalkyl Substances and Interim Guidance for Clinicians Responding to Patient Exposure Concerns Interim Guidance*. [https://www.atsdr.cdc.gov/pfas/docs/pfas\\_clinician\\_fact\\_sheet\\_508.pdf](https://www.atsdr.cdc.gov/pfas/docs/pfas_clinician_fact_sheet_508.pdf)
- Banzhaf, S., Filipovic, M., Lewis, J., Sparrenbom, C. J., & Barthel, R. (2017). A review of contamination of surface-, ground-, and drinking water in Sweden by perfluoroalkyl and polyfluoroalkyl substances (PFASs). *Ambio*, *46*(3), 335–346. <https://doi.org/10.1007/s13280-016-0848-8>
- 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. (2017). Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater. *Environmental Science and Technology*, *51*(4), 2047–2057. <https://doi.org/10.1021/acs.est.6b05843>
- Beeson, S., & Martin, J. W. (2015). Isomer-specific binding affinity of perfluorooctanesulfonate (PFOS) and perfluorooctanoate (PFOA) to serum proteins. *Environmental Science and Technology*, *49*(9), 5722–5731. <https://doi.org/10.1021/es505399w>
- Bhattacharya, A. A., Grüne, T., & Curry, S. (2000). Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *Journal of Molecular Biology*, *303*(5), 721–732. <https://doi.org/10.1006/jmbi.2000.4158>
- Bischel, H. N. H. N., MacManus-Spencer, L. A. L. A., & Luthy, R. G. R. G. (2010). Noncovalent interactions of long-chain perfluoroalkyl acids with serum albumin. *Environmental Science and Technology*, *44*(13), 5263–5269. <https://doi.org/10.1021/es101334s>
- Bischel, H. N., Macmanus-Spencer, L. A., Zhang, C., & Luthy, R. G. (2011). Strong associations of short-chain perfluoroalkyl acids with serum albumin and investigation of binding mechanisms. *Environmental Toxicology and Chemistry*, *30*(11), 2423–2430. <https://doi.org/10.1002/etc.647>
- Buck, R. C., Franklin, J., Berger, U., Conder, J. M., Cousins, I. T., Voogt, P. de, Jensen, A. A., Kannan, K., Mabury, S. A., Pj, S., & Leeuwen, V. (2011). *Perfluoroalkyl and Polyfluoroalkyl Substances in the Environment : Terminology , Classification , and Origins*. *7*(4), 513–541. <https://doi.org/10.1002/ieam.258>
- Bugsel, B., & Zwiener, C. (2020). LC-MS screening of poly- and perfluoroalkyl substances in contaminated soil by Kendrick mass analysis. *Analytical and Bioanalytical Chemistry*, *412*(20), 4797–4805. <https://doi.org/10.1007/s00216-019-02358-0>
- Calafat, A. M., Wong, L. Y., Kuklennyik, Z., Reidy, J. A., & Needham, L. L. (2007). Polyfluoroalkyl chemicals in the U.S. population: Data from the national health and nutrition examination survey (NHANES) 2003–2004 and comparisons with NHANES 1999–2000. *Environmental Health Perspectives*, *115*(11), 1596–1602. <https://doi.org/10.1289/ehp.10598>
- Chen, Y. C. (2015). Beware of docking! *Trends in Pharmacological Sciences*, *36*(2), 78–95. <https://doi.org/10.1016/j.tips.2014.12.001>

- Chen, Y. M., & Guo, L. H. (2009). Fluorescence study on site-specific binding of perfluoroalkyl acids to human serum albumin. *Archives of Toxicology*, 83(3), 255–261. <https://doi.org/10.1007/s00204-008-0359-x>
- Chi, Q., Li, Z., Huang, J., Ma, J., & Wang, X. (2018). Interactions of perfluorooctanoic acid and perfluorooctanesulfonic acid with serum albumins by native mass spectrometry, fluorescence and molecular docking. *Chemosphere*, 198, 442–449. <https://doi.org/10.1016/j.chemosphere.2018.01.152>
- Daniels, R. D., Bertke, S., Dahm, M. M., Yiin, J. H., Kubale, T. L., Hales, T. R., Baris, D., Zahm, S. H., Beaumont, J. J., Waters, K. M., & Pinkerton, L. E. (2015). Exposure–response relationships for select cancer and non- cancer health outcomes in a cohort of US firefighters from San Francisco, Chicago and Philadelphia (1950–2009). *Occup Environ Med.*, 72(10), 699–706. <https://doi.org/10.1136/oemed-2014-102671>
- Danish Ministry of the Environment. (2015). *Short-chain Polyfluoroalkyl Substances (PFAS)* (J. Kjølholt, A. A. Jensen, & M. Warming, Eds.; Issue 1707). The Danish Environmental Protection Agency.
- Dietzen, D. J. (2018). Amino acids, peptides, and proteins. In *Principles and Applications of Molecular Diagnostics*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-816061-9.00013-8>
- Gan, J., Zhang, H., & Humphreys, W. G. (2016). Drug-Protein Adducts: Chemistry, Mechanisms of Toxicity, and Methods of Characterization. *Chemical Research in Toxicology*, 29(12), 2040–2057. <https://doi.org/10.1021/acs.chemrestox.6b00274>
- Gyllenhammar, I., Berger, U., Sundström, M., McCleaf, P., Eurén, K., Eriksson, S., Ahlgren, S., Lignell, S., Aune, M., Kotova, N., & Glynn, A. (2015). Influence of contaminated drinking water on perfluoroalkyl acid levels in human serum - A case study from Uppsala, Sweden. *Environmental Research*, 140, 673–683. <https://doi.org/10.1016/j.envres.2015.05.019>
- Hamdi, O. A. A., Feroz, S. R., Shilpi, J. A., Anouar, E. H., Mukarram, A. K., Mohamad, S. B., Tayyab, S., & Awang, K. (2015). Spectrofluorometric and molecular docking studies on the binding of curcumenol and curcumenone to human serum albumin. *International Journal of Molecular Sciences*, 16(3), 5180–5193. <https://doi.org/10.3390/ijms16035180>
- Han, X., Snow, T. A., Kemper, R. A., & Jepson, G. W. (2003). Binding of perfluorooctanoic acid to rat and human plasma proteins. *Chemical Research in Toxicology*, 16(6), 775–781. <https://doi.org/10.1021/tx034005w>
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Advances in Mathematics*, 4(17). <https://doi.org/10.1016/j.aim.2014.05.019>
- Hirs, C. H. W., Stein, W. H., & Moore, S. (1954). The Free Amino Acids of Human Blood Plasma. *Journal of Biological Chemistry*, 211, 941–950.
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M. Y., Nakanishi, H., Ikeda, K., ... Nishioka, T. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703–714. <https://doi.org/10.1002/jms.1777>
- Houtz, E. F., Higgins, C. P., Field, J. A., & Sedlak, D. L. (2013). Persistence of perfluoroalkyl acid precursors in AFFF-impacted groundwater and soil. *Environmental Science and Technology*, 47(15), 8187–8195. <https://doi.org/10.1021/es4018877>
- Kaboré, H. A., Vo Duy, S., Munoz, G., Méité, L., Desrosiers, M., Liu, J., Sory, T. K., & Sauvé, S. (2018). Worldwide drinking water occurrence and levels of newly-identified perfluoroalkyl and

- polyfluoroalkyl substances. *Science of the Total Environment*, 616–617, 1089–1100. <https://doi.org/10.1016/j.scitotenv.2017.10.210>
- Krieg, E., Weissman, H., Shimon, E., Baris, A., & Rybtchinski, B. (2014). Understanding the effect of fluorocarbons in aqueous supramolecular polymerization: Ultrastrong noncovalent binding and cooperativity. *Journal of the American Chemical Society*, 136(26), 9443–9452. <https://doi.org/10.1021/ja503906p>
- Lapierre, H., Binggeli, S., Sok, M., Pellerin, D., & Ouellet, D. R. (2019). Estimation of correction factors to determine the true amino acid concentration of protein after a 24-hour hydrolysis. *Journal of Dairy Science*, 102(2), 1205–1212. <https://doi.org/10.3168/jds.2018-15392>
- Larsen, M. T., Kuhlmann, M., Hvam, M. L., & Howard, K. A. (2016). Albumin-based drug delivery: harnessing nature to cure disease. *Molecular and Cellular Therapies*, 4(1), 1–12. <https://doi.org/10.1186/s40591-016-0048-8>
- Lau, C., Dagnino, S., Kato, K., Ye, X., Calafat, A. M., Rooney, A. A., Boyles, A. L., Walker, V. R., Wambaugh, J., Reiner, J. L., Place, B. J., Butenhoff, J. L., Rodricks, J. V., Kudo, N., Jiang, Q., Gao, H., Zhang, L., Keil, D. E., Reed, C. E., ... DeWitt, J. C. (2015). *Toxicological Effects of Perfluoroalkyl and Polyfluoroalkyl Substances* (J. C. Dewitt & R. R. Dietert, Eds.). Humana Press, Cham. <https://doi.org/10.1007/978-3-319-15518-0>
- Li, Y., Fletcher, T., Mucs, D., Scott, K., Lindh, C. H., Tallving, P., & Jakobsson, K. (2018). Half-lives of PFOS, PFHxS and PFOA after end of exposure to contaminated drinking water. *Occupational and Environmental Medicine*, 75(1), 46–51. <https://doi.org/10.1136/oemed-2017-104651>
- Lingala, S. M., & Ghany, M. G. M. Mhs. (2016). *Per- and polyfluoroalkyl substances (PFAS) blood levels after contamination of a community water supply and comparison with 2013-14 NHANES*. 25(3), 289–313. <https://doi.org/110.1016/j.bbi.2017.04.008>
- McDonough, C. A., Choyke, S., Ferguson, P. L., Dewitt, J., & Higgins, C. P. (2020). Bioaccumulation of Novel Per-and Polyfluoroalkyl Substances (PFASs) in Mice Dosed with an Aqueous Film-Forming Foam (AFFF). *Environmental Science & Technology*. <https://doi.org/10.1021/acs.est.0c00234>
- Muñoz, A., Kral, R., & Schimmel, H. (2011). Quantification of protein calibrants by amino acid analysis using isotope dilution mass spectrometry. *Analytical Biochemistry*, 408(1), 124–131. <https://doi.org/10.1016/j.ab.2010.08.037>
- Mustățea, G., Ungureanu, E. L., & Iorga, E. (2019). Protein acidic hydrolysis for amino acids analysis in food - progress over time: A short review. *Journal of Hygienic Engineering and Design*, 26(April), 81–87.
- Myatt, D. P. (2017). The correlation of plasma proteins binding capacity and flavopiridol cellular and clinical trial studies. *Biomedical Spectroscopy and Imaging*, 6(1–2), 59–73. <https://doi.org/10.3233/bsi-170165>
- Nakayama, S. F., Yoshikane, M., Onoda, Y., Nishihama, Y., Iwai-Shimada, M., Takagi, M., Kobayashi, Y., & Isobe, T. (2019). Worldwide trends in tracing poly- and perfluoroalkyl substances (PFAS) in the environment. *TrAC - Trends in Analytical Chemistry*, 121, 115410. <https://doi.org/10.1016/j.trac.2019.02.011>
- Ng, C. A., & Hungerbuehler, K. (2015). Exploring the Use of Molecular Docking to Identify Bioaccumulative Perfluorinated Alkyl Acids (PFAAs). *Environmental Science and Technology*, 49(20), 12306–12314. <https://doi.org/10.1021/acs.est.5b03000>
- Ng, C. A., & Hungerbuehler, K. (2014). Bioaccumulation of perfluorinated alkyl acids: Observations and models. *Environmental Science and Technology*, 48(9), 4637–4648. <https://doi.org/10.1021/es404008g>

- Nielsen, C. J. (Norwegian E. A. (2017). *Potential PFBS and PFHxS Precursors-Literature study on abiotic degradation processes of abiotic degradation pathways leading to PFBS and PFHxS*.
- Numata, J., Kowalczyk, J., Adolphs, J., Ehlers, S., Schafft, H., Fuerst, P., Müller-Graf, C., Lahrsen-Wiederholt, M., & Greiner, M. (2014). Toxicokinetics of seven perfluoroalkyl sulfonic and carboxylic acids in pigs Fed a contaminated diet. *Journal of Agricultural and Food Chemistry*, 62(28), 6861–6870. <https://doi.org/10.1021/jf405827u>
- OECD. (2018). Toward a new comprehensive global database of per- and polyfluoroalkyl substances (PFASs): Summary Report on Updating the OECD 2007 List of Per- and Polyfluoroalkyl Substances (PFASs). *Series on Risk Management*, 39, 1–24.
- Olsen, G. W., Burris, J. M., Ehresman, D. J., Froelich, J. W., Seacat, A. M., Butenhoff, J. L., & Zobel, L. R. (2007). Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. *Environmental Health Perspectives*, 115(9), 1298–1305. <https://doi.org/10.1289/ehp.10009>
- Otter, D. E. (2012). Standardised methods for amino acid analysis of food. *British Journal of Nutrition*, 108(SUPPL. 2), 230–237. <https://doi.org/10.1017/S0007114512002486>
- Pei, Y., Li, H., & You, J. (2017). Determining equilibrium partition coefficients between lipid/protein and polydimethylsiloxane for highly hydrophobic organic contaminants using preloaded disks. *Science of the Total Environment*, 598, 385–392. <https://doi.org/10.1016/j.scitotenv.2017.04.123>
- Place, B. J., & Field, J. A. (2012). Identification of novel fluorochemicals in aqueous film-forming foams used by the US military. *Environmental Science and Technology*, 46(13), 7120–7127. <https://doi.org/10.1021/es301465n>
- Rand, A. A., & Mabury, S. A. (2012). In vitro interactions of biological nucleophiles with fluorotelomer unsaturated acids and aldehydes: Fate and consequences. *Environmental Science and Technology*, 46(13), 7398–7406. <https://doi.org/10.1021/es3008485>
- Rand, A. A., & Mabury, S. A. (2013). Covalent binding of fluorotelomer unsaturated aldehydes (FTUALs) and carboxylic acids (FTUCAs) to proteins. *Environmental Science and Technology*, 47(3), 1655–1663. <https://doi.org/10.1021/es303760u>
- Rotander, A., Kärrman, A., Toms, L. M. L., Kay, M., Mueller, J. F., & Gómez Ramos, M. J. (2015). Novel fluorinated surfactants tentatively identified in firefighters using liquid chromatography quadrupole time-of-flight tandem mass spectrometry and a case-control approach. *Environmental Science and Technology*, 49(4), 2434–2442. <https://doi.org/10.1021/es503653n>
- Salvalaglio, M., Musciconico, I., & Cavallotti, C. (2010). Determination of energies and sites of binding of PFOA and PFOS to human serum albumin. *Journal of Physical Chemistry B*, 114(46), 14860–14874. <https://doi.org/10.1021/jp106584b>
- Sander, T., Freyss, J., von Korff, M., & Rufener, C. (2015). DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling*, 55(2), 460–473. <https://doi.org/10.1021/ci500588j>
- Schrodinger LLC. (2015a). *The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8*.
- Schrodinger LLC. (2015b). *The PyMOL Molecular Graphics System, Version 1.8*.
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology*, 48(4), 2097–2098. <https://doi.org/10.1021/es5002105>
- Schymanski, E. L., Singer, H. P., Longrée, P., Loos, M., Ruff, M., Stravs, M. A., Ripollés Vidal, C., & Hollender, J. (2014). Strategies to characterize polar organic contamination in wastewater:

- Exploring the capability of high resolution mass spectrometry. *Environmental Science and Technology*, 48(3), 1811–1818. <https://doi.org/10.1021/es4044374>
- Sobolewski, K., Radparvar, S., Wong, C., & Johnston, J. (2018). Blood, Blood Components, Plasma, and Plasma Products. *Side Effects of Drugs Annual*, 40, 415–429. <https://doi.org/10.1016/bs.seda.2018.06.011>
- Sunderland, E. M., Hu, X. C., Dassuncao, C., Tokranov, A. K., Wagner, C. C., & Allen, J. G. (2019). A Review of the Pathways of Human Exposure to Poly- and Perfluoroalkyl Substances (PFASs) and Present Understanding of Health Effects. *Journal of Exposure Science & Environmental Epidemiology*, 29(2), 131–147. <https://doi.org/10.1038/s41370-018-0094-1>
- Trott, O., & Olson, A. (2010). Autodock vina: improving the speed and accuracy of docking. *Journal of Computational Chemistry*, 31(2), 455–461. <https://doi.org/10.1002/jcc.21334>
- Trowbridge, J., Gerona, R. R., Lin, T., Rudel, R. A., Bessonneau, V., Buren, H., & Morello-Frosch, R. (2020). Exposure to Perfluoroalkyl Substances in a Cohort of Women Firefighters and Office Workers in San Francisco. *Environmental Science and Technology*, 54(6), 3363–3374. <https://doi.org/10.1021/acs.est.9b05490>
- Utrecht, J. (2009). *Immune-Mediated Adverse Drug Reactions*. 24–34.
- vanden Heuvel, J. P., Kuslikis, B. I., & Peterson, R. E. (1992). Covalent Binding of Perfluorinated Fatty Acids to Proteins in the Plasma, Liver and Testes of Rats. *Chemical-Biological Interactions*, 82, 317–328.
- Wang, Z., Dewitt, J. C., Higgins, C. P., & Cousins, I. T. (2017). A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environmental Science and Technology*, 51(5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>
- Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S., & Hou, T. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics*, 18(18), 12964–12975. <https://doi.org/10.1039/c6cp01555g>
- Wardell, M., Wang, Z., Ho, J. X., Robert, J., Ruker, F., Ruble, J., & Carter, D. C. (2002). The atomic structure of human methemalbumin at 1.9 Å. *Biochemical and Biophysical Research Communications*, 291(4), 813–819. <https://doi.org/10.1006/bbrc.2002.6540>
- Wilbur, D. S. (2003). *Formation of Sulfonamide Bonds Through Reaction of Dyes with Serum Proteins Correlation of Tumor Radiation-Absorbed Dose with Response Is Easier to Find in Previously Untreated Patients*. 44(9), 1540–1545.
- Woodcroft, M. W., Ellis, D. A., Rafferty, S. P., Burns, D. C., March, R. E., Stock, N. L., Trumpour, K. S., Yee, J., & Munro, K. (2010). Experimental characterization of the mechanism of perfluorocarboxylic acids' liver protein bioaccumulation: The key role of the neutral species. *Environmental Toxicology and Chemistry*, 29(8), 1669–1677. <https://doi.org/10.1002/etc.199>
- Wu, L.-L., Gao, H.-W., Gao, N.-Y., Chen, F.-F., & Chen, L. (2009). Interaction of perfluorooctanoic acid with human serum albumin. *BMC Structural Biology*, 9(1), 31. <https://doi.org/10.1186/1472-6807-9-31>
- Yang, F., Zhang, Y., & Liang, H. (2014). Interactive association of drugs binding to human serum albumin. *International Journal of Molecular Sciences*, 15(3), 3580–3595. <https://doi.org/10.3390/ijms15033580>
- Zhang, L., Ren, X. M., & Guo, L. H. (2013). Structure-based investigation on the interaction of perfluorinated compounds with human liver fatty acid binding protein. *Environmental Science and Technology*, 47(19), 11293–11301. <https://doi.org/10.1021/es4026722>

Zhang, X., Chen, L., Fei, X. C., Ma, Y. S., & Gao, H. W. (2009). Binding of PFOS to serum albumin and DNA: Insight into the molecular toxicity of perfluorochemicals. *BMC Molecular Biology*, *10*, 1–12. <https://doi.org/10.1186/1471-2199-10-16>

## CHAPTER 4: DEVELOPMENT AND PREDICTION OF PER- AND POLYFLUOROALKYL SUBSTANCES (PFAS) TOXICITY USING MOLECULAR DOCKING AND QSAR TOOLS

### 4.1 Introduction

Poly- and perfluoroalkyl substances (PFAS) are a class of ubiquitous environmental contaminants sourced from a broad range of commercial and industrial products (Glüge et al., 2020). PFAS exposures are linked to many adverse health effects including increased cholesterol levels, changes in liver enzymes, decrease in infant birth weights, decreased vaccine response in children, increased risk of pre-eclampsia in pregnant women, and increased risk of kidney or testicular cancer (Pelch et al., 2019). Yet the majority of PFAS toxicity studies focus on only a small subset of PFAS (i.e., most often perfluoroalkyl acids, PFAAs). A broader range of adverse health effects are expected for the diverse structures that make up different PFAS sub-classes (Allendorf et al., 2019). Additionally, many regulatory entities have adopted a new, broader definition of PFAS by the Organisation for Economic Co-operation and Development (OECD) (Wang et al., 2021). The new definition removed the aliphatic requirement defined by Buck et al. (2011). As a result, the number of PFAS traded in the global market and present in the environment has increased from ~3,000 estimated by Wang et al. (2017) to over six million (<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120>; OECD PFAS tree published on June 10<sup>th</sup> 2022). Understanding toxicity risks for all known PFAS is critical for developing appropriate and coordinated chemicals management.

Conducting chemical risk assessments for all manufactured PFAS using *in vivo* or *in vitro* approaches is too burdensome when considering the time, instrument, material, and labor costs. Moreover, commercial standards or analytical methods remain unavailable for the majority of

PFAS (Liu et al., 2019). Computational approaches offer robust and high-throughput alternatives to uncover structural-toxicity relationships. Read-across approaches that involve construction of toxicological prediction models based on quantitative structure-activity relationships (QSAR) can help address the need for screening-based risk assessments of diverse chemical classes. QSAR has been widely adopted by regulatory agencies (e.g., OECD (OECD, 2012), REACH (European Chemicals Agency, 2016) and US EPA (Environmental Protection Agency & Todd, 2020)) to assess ecotoxicity and health hazards of pesticides (Hamadache et al., 2017). Cheng and Ng (2019) developed the first machine learning (ML) based QSAR model for PFAS. They identified multitask neural network to be the best performing machine learning model, which was used to classify the bioactivities of 3,486 PFAS based on 26 bioactivities of 1,012 molecules with  $-C_3F_6-$  moiety. Interestingly, majority of the bioactive PFAS contained less than 12 perfluorocarbons. Chen and Ng's study was the first effort on understanding PFAS health impact using data science. However, considering PFAS definition has changed since then, an updated training data for QSAR model is needed. More importantly, most of the training data used in Chen and Ng's study was obtained from pharmaceutical bioassays (i.e., testing on cancer cells), a direct correlation between the model prediction and negative health impact is infeasible.

I postulate that QSAR-based predictive models for adverse health impacts can be improved by incorporating information on PFAS-protein interactions. Adverse health outcomes induced by PFAS (e.g., liver toxicity, lipid and insulin dysregulation, and cancer) have been linked to interactions with receptor and carrier proteins. Strong binding to nuclear receptors, especially peroxisome proliferator activated receptors (PPARs), was identified as one potential pathway for PFAS-induced toxicity (A. Behr et al., 2020; Fenton et al., 2021; Kirk et al., 2021). PPARs ( $-\alpha$ ,  $-\beta/\delta$ , and  $-\gamma$ ) belong to a ligand-activated nuclear hormone receptor superfamily whose ligands

include steroids, thyroid hormone, retinoic acid, and vitamin D. PPARs control lipid metabolism and catabolism (fatty acid transport, cell uptake, intracellular binding, and activation of storage) (Tyagi et al., 2011). Liver- and intestinal fatty acid binding proteins (L-FABP and I-FABP) are another important class of proteins involved in regulating the toxicokinetics of PFAS (Woodcroft et al., 2010; Yang et al., 2020; Zhang et al., 2013). In addition to facilitating fatty-acid transportation and accumulation, FABPs deliver lipophilic ligands to PPARs (Velkov, 2013; Wolfrum et al., 2001). Human serum albumin (HSA), the most abundant blood protein, is another important protein thought to play a role in PFAS body burden and pharmacokinetics. Strong binding of some PFAS to HSA (Allendorf et al., 2019; Bischel et al., 2011; Li et al., 2021), for instance, may contribute to the extended half-lives of PFAS in humans (i.e., legacy PFAS has serum elimination half-life of 3.5 to 8.5 years) (Olsen et al., 2007). In combination with active renal tubular reabsorption, bioaccumulative PFAS render prolonged bioactivity/toxicity (Ferrari et al., 2019).

In this study, I aim to build a multi-condition QSAR model for PFAS toxicity/bioactivity (BioTox) prediction and screening that incorporates associations of PFAS with PPARs, FABPs, and HSA. I use an *in silico* approach to characterize the PFAS-protein interactions, such as molecular docking, which have already been used successfully to predict binding affinities of diverse PFAS with each of these protein targets (Cheng & Ng, 2018; Chi et al., 2018; Li et al., 2021; Ng & Hungerbuehler, 2015). I build on these computational pipelines to expand the number of PFAS analyzed, and I integrate results with QSAR to enhance the model's predictive power for toxicity or bioactivity endpoints. My overall approach includes (1) curating a PFAS toxicity database and using toxicity or human health relevance bioactivity data as a training set, (2) comparing and validating linear and nonlinear models (six different ML tools), (3) exploring the

significance of HSA, FABP, PPAR docking scores amongst other physico-chemical descriptors in toxicity predictions, and (4) predicting toxicity/bioactivity of PFAS that are known to be manufactured and in use.

## 4.2 Method

### 4.2.1 Computational resource and machine learning packages

All computational simulations were run on a Linux machine with 11th Generation Intel® Xeon® CPU and the graphic based simulations (e.g., molecular docking and molecular dynamics) were accelerated with a Nvidia Quadro RTX 5000 GPU (16GB, 4DP Precision 3650T).

### 4.2.2 Chemical database curation

First, I downloaded 4,760 PFAS on OECD PFAS manufactured and used based on PubChem classification ([PubChem \(nih.gov\)](https://pubchem.ncbi.nlm.nih.gov) accessed on July 8<sup>th</sup> 2022). Second, I generated 3D conformers using smiles code with “random, low energy bias” algorithm and FFFF94s+forcefield algorithm in DataWarrior 5.5.0 (López-López et al., 2019). Third, I blind docked 4,390 structures to HSA crystal structure (PDB: 1E7G), PPAR- $\alpha$ , - $\beta/\delta$ , and  $\gamma$  crystal structures (PDB: 4CI4, 3U9Q, and 3TKM) and L-FABP and I-FABP crystal structures (PDB: 3STM and 3AKM using Docking App RF (Macari et al., 2020), which uses Autodock Vina’s (Ahmed ElTijani, 2019) searching algorithm and improved scoring function for improved predictions. I successfully docked 4,098 compounds to all receptors (repeated 10 times). Docking failure was caused primarily by a lack of atomic parameter files for certain atoms (e.g., rhodium, chromium, tin, silicon etc.) in some PFAS. Given the docking distribution and the known fatty-acid (FA) binding sites on HSA, we sub-characterized the binding pockets of HSA into six binding pockets (e.g., enlarged grid box size by 25 Å<sup>2</sup> compared to Li et al. (2021)). The minimum docking scores of each binding pocket and physico-chemical properties simulated using DataWarrior 5.5.0 are listed in Table 2.

Table 1. List of descriptors used in development of QSAR model

1	Min docking score FA1 (kcal/mol)	29	Stereo Centers
2	Min docking score FA2 (kcal/mol)	30	Rotatable Bonds
3	Min docking score FA34 (kcal/mol)	31	Rings Closures
4	Min docking score FA5 (kcal/mol)	32	Aromatic Atoms
5	Min docking score FA6 (kcal/mol)	33	sp <sup>3</sup> -Atoms
6	Min docking score FA7 (kcal/mol)	34	Symmetric atoms
7	Min docking score IFABP (kcal/mol)	35	Small Rings
8	Min docking score LFABP (kcal/mol)	36	Carbo-Rings
9	Min docking score PPAR $\alpha$ (kcal/mol)	37	Hetero-Rings
10	Min docking score PPAR $\beta$ (kcal/mol)	38	Saturated Rings
11	Min docking score PPAR $\gamma$ (kcal/mol)	39	Non-Aromatic Rings
12	Monoisotopic Mass	40	Aromatic Rings
13	cLogP	41	Saturated Carbo-Rings
14	cLogS	42	Non-Aromatic Carbo-Rings
15	H-Acceptors	43	Carbo-Aromatic Rings
16	H-Donors	44	Saturated Hetero-Rings
17	Total Surface Area	45	Non-Aromatic Hetero-Rings
18	Relative PSA	46	Hetero-Aromatic Rings
19	Polar Surface Area	47	Amides
20	Druglikeness	48	Amines
21	Shape Index	49	Alkyl-Amines
22	Molecular Flexibility	50	Aromatic Amines
23	Molecular Complexity	51	Aromatic Nitrogens
24	Fragments	52	Basic Nitrogens
25	Non-H Atoms	53	Acidic Oxygens
26	Non-C/H Atoms	54	Globularity SVD
27	Metal-Atoms	55	VDW-Surface
28	Electronegative Atoms	56	VDW-Volume

#### 4.2.3 Bioactivity database curation

The toxic/bioactive dataset was constructed by searching -C<sub>2</sub>F<sub>4</sub>- substructures from public databases: Tox21- qualitative *in vitro* toxicity measurements on 12 biological targets, including nuclear receptors and stress response pathways (Tice et al., 2013); ToxCast<sup>TM</sup> - *in vitro* high-throughput screening, including experiments on over 600 tasks; SIDER - database of marketed drugs and adverse drug reactions (ADR) (Kuhn et al., 2016); ClinTox - qualitative data of drugs

approved by the (Food and Drug Administration) FDA and those that have failed clinical trials for toxicity reasons; and acute toxicity data for median lethal dose (LD<sub>50</sub>) archived from National Toxicology Program (NTP, 2006). 112 chemicals in different databases overlap, 23 of which have conflicting information (e.g., toxic in one database but not toxic in another database). For these chemicals, I selected *in vivo* data over *in vitro* data (e.g., Acute rat oral LD<sub>50</sub> by NTP vs Tox21).

All chemicals are classified in a binary fashion (e.g., toxic vs nontoxic or bioactive vs bio-inactive). For Tox21 and ToxCast, any active assay (e.g., reporter gene activation) grants the chemical to be classified as a bioactive compound. For SIDER, I classify chemicals with over 14 organ-related adverse effects out of 27 total side effects as bioactive compounds. For ClinTox data, I consider any drug pose toxicological effect as toxic even if it was FDA approved. For acute rat oral LD<sub>50</sub>, I follow World Health Organization (WHO) acute hazard rankings. If the value is under 2,000 mg/kg the chemical is considered at least moderately hazardous (a value of 5-50 mg/kg is highly hazardous, and values <5 mg/kg are extremely hazardous chemicals). Overall, we curated 796 toxicity/bioactivity datapoints (Table 1). While some of the bioassay data used to construct BioTox model is not typically use to define toxicity, PFAS binding to these critical receptors or causing large number of biological side effects is highly concerning.

Table 2. Summary of the five toxicity/bioactivity datasets used to build QSAR model.

Database	Total number of PFAS	% Toxic
Tox21 by EPA	251	47%
ToxCast by EPA	282	36%
SIDER	71	78%
ClinTox by FDA	72	15%
Acute rat oral LD <sub>50</sub> by NTP	120	73%

#### 4.2.4 QSAR model tools

I use *QSAR-Co-X*, an open-source python-based application developed by Halder & Dias Soeiro Cordeiro (2021) to perform (1) Sequential Forward Selection-Linear Discrimination Analysis (SFS-LDA), a linear model which achieves better reproducibility in comparison to Genetic Algorithms (GA)-LDA (Halder & Cordeiro, 2019). (2) a suite of nonlinear machine learning models ( $k$ -Nearest Neighbors (kNN), Support Vector Classifier (SVC), Random Forests (RF), Bernoulli Naïve Bayes classifier (NB), Gradient Boosting (GB), and Multilayer Perceptron (MLP) neural networks). The SFS-LDA model is employed to investigate the significance of chemical descriptors effect on toxicity.

There are three major steps to build a QSAR model. First, the dataset is split into training (80%;  $N_{\text{Train}} = 636$ ) and validation (20%;  $N_{\text{val}} = 159$ ) sets using the random forest division method. Then I apply the simple Box-Jenkins (Casañola-Martin et al., 2015) moving average for all the descriptors in the training sets with different toxicology end-points (and experimental conditions). In brief, modified descriptors are computed by taking the difference between the input descriptors of chemicals and their arithmetic mean for a specific condition (i.e., bioactivity data is defined in two conditions: source of data and type of bioassay). The training set and their modified descriptors are further divided into sub-training and testing sets used to model optimization. The descriptors in validation set are externally modified in the same fashion to evaluate the occurrence of model overfitting.

#### 4.2.5 Model optimization and validation

I use several performance metrics to evaluate the performance of all the machine learning and linear models. These include: area under the curve (AUC) of the receiver operating characteristic (ROC) curve (Hanczar et al., 2010), Matthews Correlation Coefficient (MCC)

(Boughorbel et al., 2017), accuracy, F1 score, specificity, and sensitivity (Jain & Nicholls, 2008).

I calculate these metrics (Equation 6 to 11) based on prediction outcomes in discrimination analysis, which consists of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Note: the rate of each term is denoted in lower case (e.g.,  $\frac{tp}{tp+fn}$  is the rate of sensitivity for true positive).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (7)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$\text{F1 score} = \frac{2*TP}{2*TP+FN+FP} \quad (9)$$

$$\text{MCC} = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (10)$$

$$\text{AUC-ROC} = \int_0^1 \frac{tp}{tp+fn} d\left(\frac{fn}{tn+fp}\right) \quad (11)$$

SFS-LDA model optimization is conducted in two steps. First, I manually changed the number of maximum features (3 to 10) while monitoring the AUC-ROC scores, which plateaued after five features were included. Second, I shuffled the response variables under experimental conditions (e.g., different bioassays) randomly for 100 times to generate 100 randomized models and compared the randomized models to the original model. This test checks if the performance of the original linear model is made by chance or true correlation. After variable responses are randomly assigned to the descriptors, true predictivity of the model is shown by comparing the statistics of the models. The original model outperformed the randomized models in accuracy (original: 93% vs random: 58%) and Wilks' lambda (original: 0.34 vs random: 0.99) (Wilks, 1932). Wilks' lambda describes how independent each input variable contributes to the model, which ranges from zero (complete discrimination) to one (no discrimination). For ML models, we used a grid search

optimization process to tune the hyperparameters based on AUC-ROC scores. The optimized hyperparameters are available in Table C1.

## 4.3 Result and discussion

### 4.3.1 Identification of significant chemical descriptors

The best SFS-LDA model for predicting toxicity/bioactivity (BioTox) was developed with five descriptors modified by the moving average of experimental conditions: C1 (database) and C2 (bioassay) in subscript shown below.

$$\text{BioTox} = 0.1344 * \text{FA1}_{(C1)} - 0.1452 * \text{FA34}_{(C1)} + 0.9308 * \text{Saturated Hetero Rings}_{(C1)} - 0.3048 * \text{PPAR } \gamma_{(C2)} - 15.51 * \text{Molecular Complexity}_{(C2)} + 4.079 \quad \dots(12)$$

The diagnostic accuracy of the model based on the AUC-ROC scores of sub-training set (0.9238), testing set (0.9080), and validation set (0.9343), is considered outstanding according to Sanders (2012) (Figure 12). The predicted toxicity/bioactivity were negatively correlated to the minimal docking scores to HSA FA3/4, PPAR  $\gamma$ , and Molecular Complexity and positively correlated to the minimal docking scores to HSA FA1 and the number of saturated heterocyclic rings.

Overall, I found PFAS-protein interactions to be important for PFAS toxicity/bioactivity predictions. Molecular docking scores for two HSA binding sites and one PPAR  $\gamma$  binding site were significantly correlated to PFAS toxicity/bioactivity prediction. Among 56 descriptors used to characterize toxicity/bioactivity of PFAS only 20% were protein docking scores, but 60% of the independent descriptors needed to construct a best SFS-LDA model were protein docking scores. As we previously reported (Li et al., 2021) and Chen (2015) discussed, molecular docking with repetitive binding protocols is a promising tool for initial screening of protein binding versus nonbinding ligands. However, weak correlations of docking results with experimental data confirmed that the approach was insufficient to differentiate strong versus weak binding chemicals

(e.g., PFAS), a weakness attributed to the lack of solvent-solvent interactions in molecular docking (Y. C. Chen, 2015). Many studies improved their prediction accuracy by using molecular dynamics (Cheng & Ng, 2018; Salvalaglio et al., 2010; Willemsen & Bourg, 2021), or flexible docking protocols inspired by molecular dynamics (Singam et al., 2020; Söderström et al., 2022). Molecular dynamics simulations are extremely time consuming. The estimated time for flexible docking of the 796 PFAS we used to build the QSAR model is 5 months (using our computational resources), while the estimated time for molecular dynamics is 4.5 years against one protein. PFAS are known to bind with multiple carrier and receptor proteins causing various adverse health outcomes (Houck et al., 2021). Although, molecular dynamics-based models can predict protein binding affinities more accurately, the approach is too time consuming to be used for predicting the toxicity of large numbers of PFAS. It is therefore encouraging that we found the more simplistic simulations to generate molecular docking scores to aid in quantitative toxicity/bioactivity predictions.

Overall bioactivity/toxicity was positively correlated to the docking predicted binding affinities to HSA FA 3/4 and PPAR  $\gamma$  (or equivalently, negative correlated to the binding scores in Kcal/mol). HSA FA 3/4 overlaps with Sudlow's binding site II, which binds with endogenous (i.e., thyroxine) and exogenous (i.e., ibuprofen) ligands. Many studies reported that PFAS preferentially bind to Sudlow's binding site II of HSA (e.g., PFBS, PFHA, PFBS, PFDoA, etc.) using fluorescence quenching or nuclear magnetic resonance (NMR) methods (Y. M. Chen & Guo, 2009; Chi et al., 2018; D'Eon et al., 2010). In an agreement with these studies, when PFAS bind to HSA Sudlow's II binding, HSA cannot transfer drugs nor perform its normal biological function (e.g., thyroid disruption). Interestingly, most studies attributed PFAS toxicity to the molecular initiation event(s) (MIE) of PPAR  $\alpha$  binding as it is activated by PFOS and PFOA at a greater

extent than PPAR  $\gamma$  (A. C. Behr et al., 2020; vanden Heuvel et al., 2006). Recent toxicology studies reveal that PFAS have multiple MIE, and PPAR  $\gamma$  activation by PFAS is linked to additional adverse health outcomes that cannot be explained by PPAR  $\alpha$  functions, such as metabolic disorder and weakness in bones (Evans et al., 2022; Kirk et al., 2021).

Unexpectedly, bioactivity/toxicity was negatively correlated to the docking-predicted binding affinities for HSA FA1. One hypothesis is that when PFAS are bound to HSA, they have less chance to interact with MIE receptors leading to lower toxicity effect. HSA has been used collectively with drugs therapeutically to reduce toxicity, collateral damage to healthy cells, and diminish the amounts of free radicals (Otagiri & Giam Chuang, 2016). Simard et al. (2006) identified FA1 as a low binding affinity site for fatty acids using NMR competition analysis, which may allow PFAS to outcompete fatty acids. In comparison to FA3/4 (or Sudlow II site), FA1 may have higher chance to cage PFAS from interacting with other MIE receptors. In this study, all fatty-acids are deleted from HSA crystal structure prior to docking. The interactions among fatty-acids, HSA, and PFAS are unknown. In future study, competitive molecular dynamics for an array of PFAS with different bioactivities could be used to test this hypothesis.

It is not surprising that PFAS bioactivity/toxicity was positively correlated to the number of heterocyclic rings and negatively correlated to molecular complexity (equation 12). Heterocyclic rings are widely found in a broad array of natural products (e.g., vitamins, hormones, and antibiotics) and are used as framework to enhance bioactivities for anthropogenic chemicals, such as pesticides (Fattorusso & Tagliatela-Scafati, 2008). Molecular complexity is calculated by taking the logarithm of the number of distinct fragments over the number of rotatable bonds. In other words, molecules with many symmetrical structures and re-occurring substructures are low complexity. By reviewing the PFAS structures with low molecular complexity, we found that most

of the toxic/bioactive ones are high in symmetry, which are one of the common causes of off-target toxic effects in drug design (Mezey, 2008).

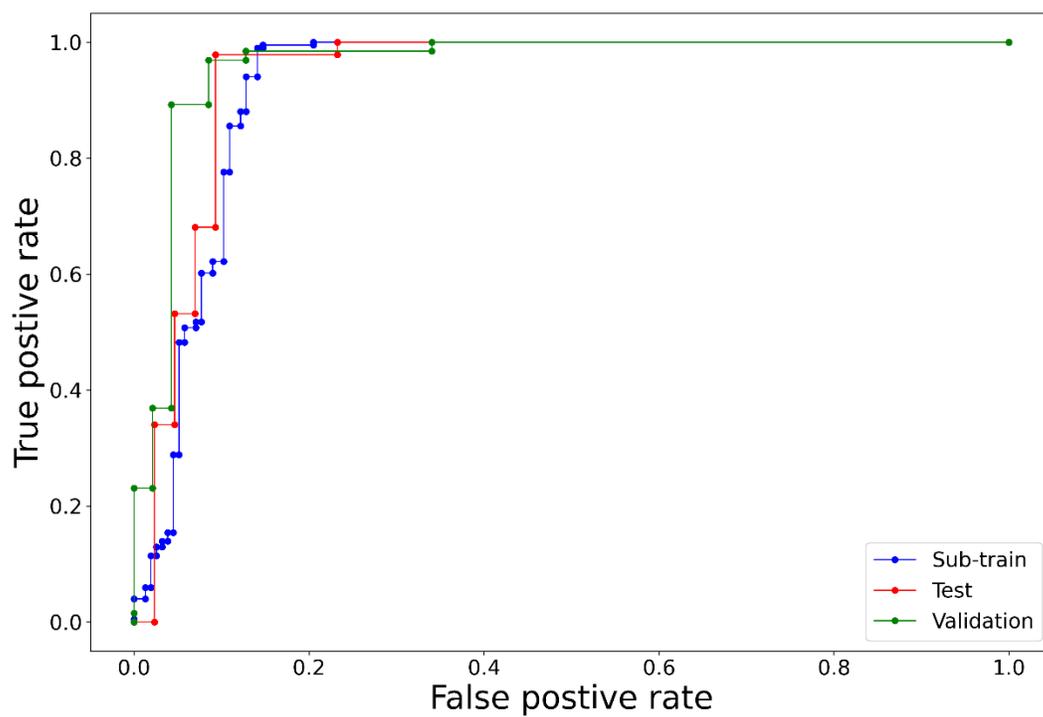
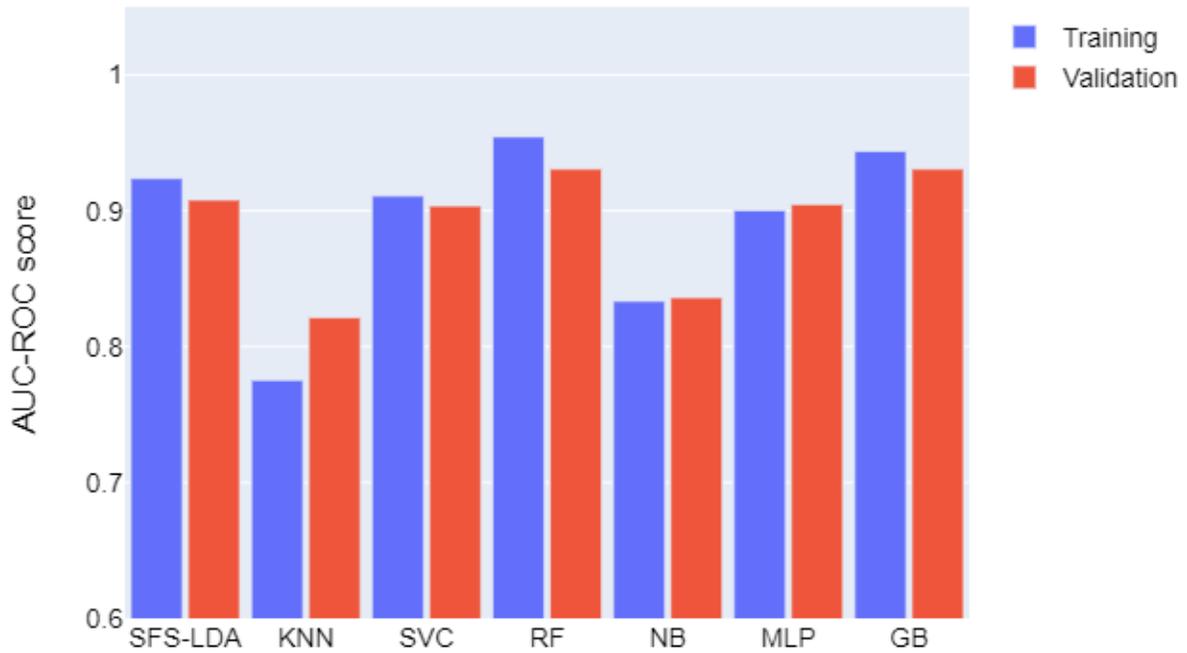


Figure 12. Receiver Operating Characteristic (ROC) curve of SFS-LDA.

### 4.3.2 Performance of QSAR models

Both RF and GB based models demonstrated the best performance based on AUC-ROC (Figure 13a) and MCC (Figure 13b) evaluation of the validation dataset. The performance of the SFS-LDA model follows closely after these top two. The AUC-ROC scores for the training set of our ML models ( $>0.9$  except NB and  $k$ NN) are slightly lower than Chen and Ng's ( $>0.95$  for all tested models), but the scores for the validation set of our ML models ( $>0.9$  except NB and  $k$ NN) are much higher than Chen and Ng's ( $\sim 0.75$ ). Unlike the model constructed by Cheng & Ng (2019) for the  $C_3F_6$  dataset, we did not use Bayesian optimization. Bayesian optimization is a powerful tuning tool for machine learning models. However, the performance metrics improvements on the training and testing sets may not translate to the external validation sets, especially when tuning models trained on small datasets (Simard et al., 2006). Indeed, an overfitting phenomenon was observed in Chen and Ng's study across the board. Since MCC values (a summary statistic that considers the unbalanced classes of the dataset) (Boughorbel et al., 2017) also agreed with AUC-ROC model, we selected the RF model to predict BioTox scores for the OECD PFAS in PubChem. Out of 4,098 manufactured and used OECD PFAS in PubChem (all successfully docked), only 2,649 PFAS are within the application domain. Here due to the limit of space, I only present RF predicted BioTox scores of 56 PFAS with a posterior probability greater than 95% for the classification in Table C2. Additional statistical metric evaluations for all tested models are available in section Figure C7.

a.



b.

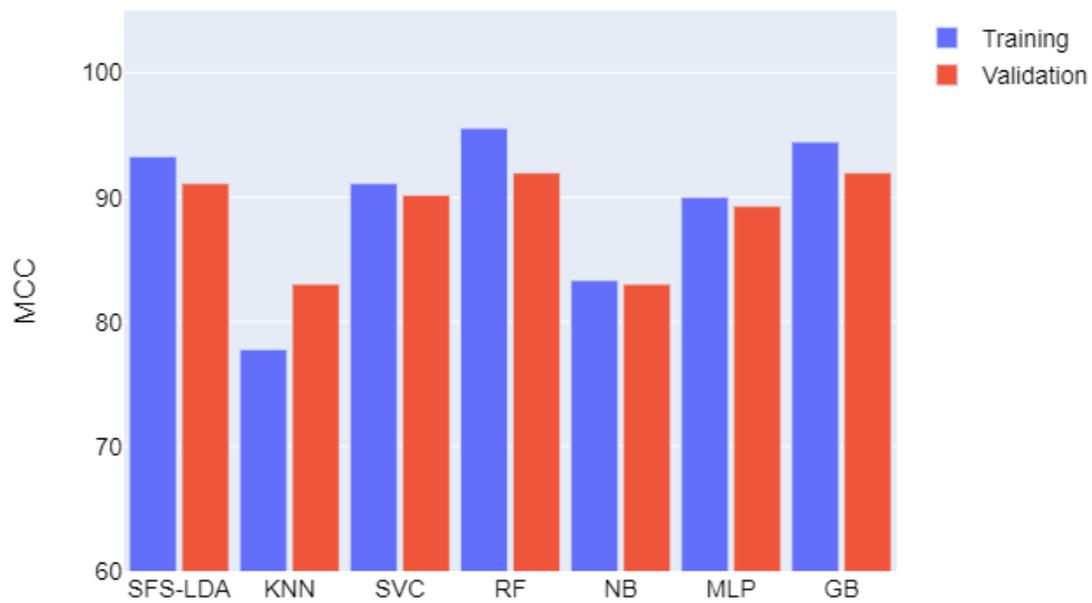


Figure 13. Predicting performance of training and validation datasets in a. Area Under The Curve - Receiver Operating Characteristics (AUC-ROC) scores and b. Matthews Correlation Coefficient (MCC) of different testing models.

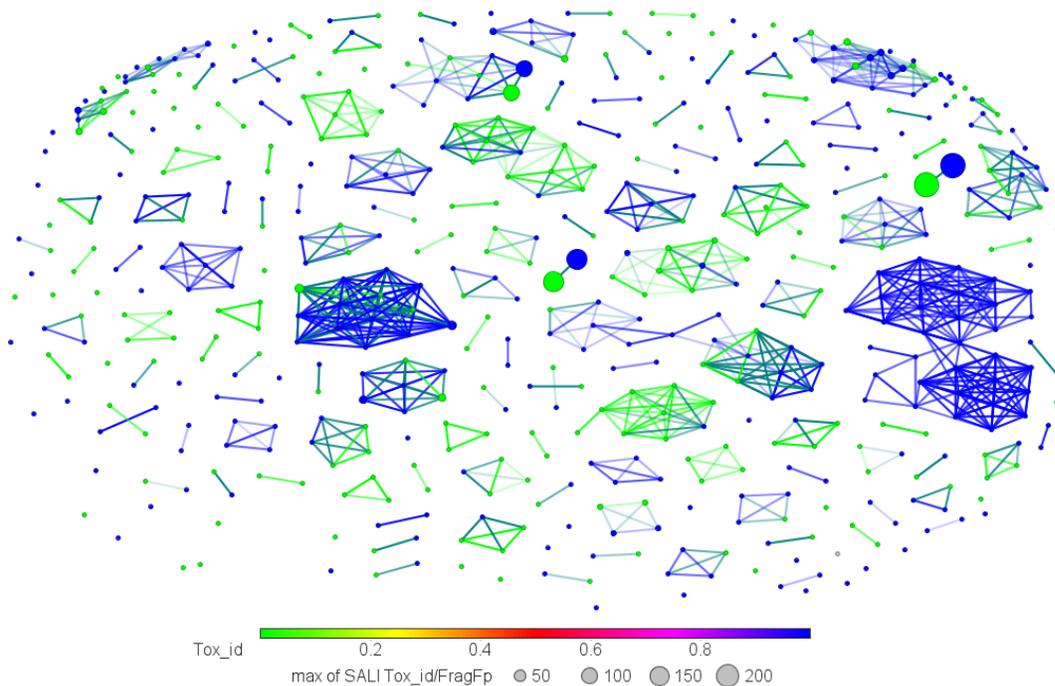
### 4.3.3 Correlating PFAS structural analysis with bioactivity/toxicity classification

In order to understand the structural similarities of active/inactive PFAS for toxicity/bioactivities, we used Structure-Activity Landscape Index (SALI) activity cliff analysis (ACA) for QSAR input data. The QSAR input data was constructed using the DataWarrior package. The SALI-ACA map (Figure 14 a) randomly positions all molecules on the 2D space and relocates the molecules based on how much activity is gained (or lost) with a small structural change. Molecules with similar structures are connected with a line and the coloring scheme are related to toxicity/bioactivity classifications (i.e., toxic/bioactive PFAS are colored blue). Most toxic/bioactive PFAS shown in blue clusters are structurally similar. Halogen substitution (e.g., containing iodine, chlorine, and bromine), carboxylic acid, benzimidazole, and dinitroaniline were common structural characteristics for these clusters. I suspect that for PFAS containing benzimidazole and dinitroaniline, the toxicity arises from the non-fluorinated moieties. Benzimidazole derivatives exhibit pharmacological activities such as antimicrobial, antiviral, anticancer, etc, and are associated with side effects and cytotoxicity (Salahuddin et al., 2017). Dinitroaniline derivatives on the other hand have been used as herbicide for decades and exhibit high toxicity to mammalian cells (J. Chen et al., 2021).

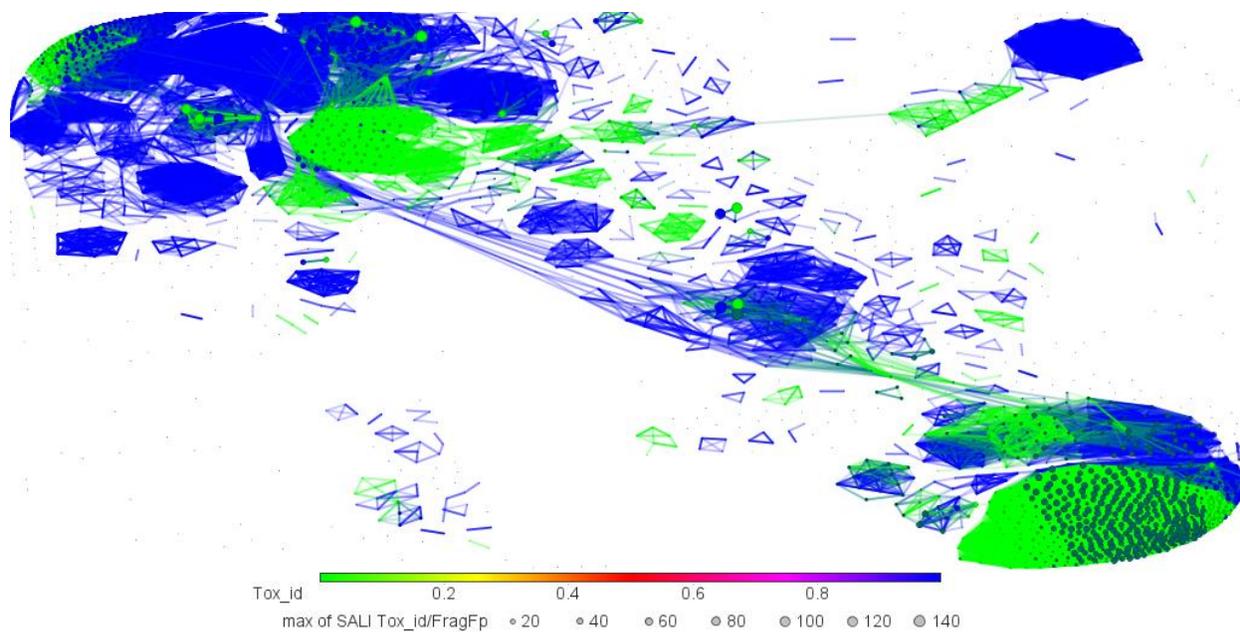
Structural variability for the large PubChem PFAS database is much greater than the input PFAS with bioactivity/toxicity data available, which poses challenges for read-across approaches. Surprisingly, such structurally diverse PFAS are within the model's application domain shown in Figure 14 (b). This might be due to the similar groups of diverse PFAS being tested and produced. However, accurate prediction of PFAS BioTox score may be enhanced for large emerging clusters rather than the ones are far from each other in the SALI-ACA map. Additionally, large numbers of activity cliffs are distributed in the large clusters merging (i.e., right bottom green cluster).

Activity cliffs are rare points representing in large bubbles, where a small change of the chemicals structures causes a large change in toxicity/bioactivity classification. PFAS in that cluster all contain long perfluorocarbon chain (more than 10 perfluorocarbons), while presenting polarize bioactivity with small changes (i.e., one halogen substitution vs hydrocarbon replacement). Altogether, *in vitro* and *in vivo* toxicity tests should, if possible, include samples of PFAS structures from each structural clusters to further inform the impacts of such activity cliffs on toxicity predictions for closely related PFAS.

a.



b.



**Figure 14.** Structure-Activity Landscape Index (SALI) activity cliff analysis (ACA) for (a) QSAR input data and (b) manufactured and used OECD PFAS in PubChem classified by bioactivity (blue: toxic/bioactive; green: nontoxic/bioinactive).

#### 4.4 Reference

- Ahmed ElTijani, M. Y. A. & A. F. A. (2019). *EasyDockVina: Graphical Interface for Ligand Optimization and High Throughput Virtual Screening with Vina (2.2)*. Zenodo. <https://doi.org/10.5281/zenodo.3732170>.
- Allendorf, F., Berger, U., Goss, K. U., & Ulrich, N. (2019). Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. *Environmental Science: Processes and Impacts*, 21(11), 1852–1863. <https://doi.org/10.1039/c9em00290a>
- Behr, A. C., Plinsch, C., Braeuning, A., & Buhrke, T. (2020). Activation of human nuclear receptors by perfluoroalkylated substances (PFAS). *Toxicology in Vitro*, 62. <https://doi.org/10.1016/j.tiv.2019.104700>
- Behr, A., Plinsch, C., Braeuning, A., & Buhrke, T. (2020). Activation of human nuclear receptors by perfluoroalkylated substances. *Toxicology in Vitro*, 62(October 2019), 104700. <https://doi.org/10.1016/j.tiv.2019.104700>
- Bischel, H. N., Macmanus-Spencer, L. A., Zhang, C., & Luthy, R. G. (2011). Strong associations of short-chain perfluoroalkyl acids with serum albumin and investigation of binding mechanisms. *Environmental Toxicology and Chemistry*, 30(11). <https://doi.org/10.1002/etc.647>
- Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE*, 12(6). <https://doi.org/10.1371/journal.pone.0177678>
- Buck, R. C., Franklin, J., Berger, U., Conder, J. M., Cousins, I. T., Voogt, P. de, Jensen, A. A., Kannan, K., Mabury, S. A., Pj, S., & Leeuwen, V. (2011). *Perfluoroalkyl and Polyfluoroalkyl Substances in the Environment: Terminology, Classification, and Origins*. 7(4), 513–541. <https://doi.org/10.1002/ieam.258>
- Casañola-Martin, G. M., Le-Thi-Thu, H., Pérez-Giménez, F., Marrero-Ponce, Y., Merino-Sanjuán, M., Abad, C., & González-Díaz, H. (2015). Multi-output model with Box–Jenkins operators of linear indices to predict multi-target inhibitors of ubiquitin–proteasome pathway. *Molecular Diversity*, 19(2), 347–356. <https://doi.org/10.1007/s11030-015-9571-9>
- Chen, J., Yu, Q., Patterson, E., Sayer, C., & Powles, S. (2021). Dinitroaniline Herbicide Resistance and Mechanisms in Weeds. In *Frontiers in Plant Science* (Vol. 12). Frontiers Media S.A. <https://doi.org/10.3389/fpls.2021.634018>
- Chen, Y. C. (2015). Beware of docking! In *Trends in Pharmacological Sciences* (Vol. 36, Issue 2, pp. 78–95). Elsevier Ltd. <https://doi.org/10.1016/j.tips.2014.12.001>
- Chen, Y. M., & Guo, L. H. (2009). Fluorescence study on site-specific binding of perfluoroalkyl acids to human serum albumin. *Archives of Toxicology*, 83(3), 255–261. <https://doi.org/10.1007/s00204-008-0359-x>
- Cheng, W., & Ng, C. A. (2018). Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach. *Environmental Science and Technology*, 52(14), 7972–7980. <https://doi.org/10.1021/acs.est.8b01268>
- Cheng, W., & Ng, C. A. (2019). Using Machine Learning to Classify Bioactivity for 3486 Per- And Polyfluoroalkyl Substances (PFASs) from the OECD List. *Environmental Science and Technology*, 53(23), 13970–13980. <https://doi.org/10.1021/acs.est.9b04833>
- Chi, Q., Li, Z., Huang, J., Ma, J., & Wang, X. (2018). Interactions of perfluorooctanoic acid and perfluorooctanesulfonic acid with serum albumins by native mass spectrometry, fluorescence

- and molecular docking. *Chemosphere*, 198, 442–449.  
<https://doi.org/10.1016/j.chemosphere.2018.01.152>
- D'Eon, J. C., Simpson, A. J., Kumar, R., Baer, A. J., & Mabury, S. A. (2010). Determining the molecular interactions of perfluorinated carboxylic acids with human sera and isolated human serum albumin using nuclear magnetic resonance spectroscopy. *Environmental Toxicology and Chemistry*, 29(8), 1678–1688. <https://doi.org/10.1002/etc.204>
- Environmental Protection Agency, & Todd. (2020). *User's Guide for T. E. S. T. (Toxicity Estimation Software Tool Version 5.1)*. <https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test>
- European Chemicals Agency. (2016). *Practical guide How to use and report (Q)SARs version 3.1*. <https://doi.org/10.2823/81818>
- Evans, N., Conley, J. M., Cardon, M., Hartig, P., Medlock-Kakaley, E., & Gray, L. E. (2022). In vitro activity of a panel of per- and polyfluoroalkyl substances (PFAS), fatty acids, and pharmaceuticals in peroxisome proliferator-activated receptor (PPAR) alpha, PPAR gamma, and estrogen receptor assays. *Toxicology and Applied Pharmacology*, 449, 116136. <https://doi.org/10.1016/j.taap.2022.116136>
- Fattorusso, Ernesto., & Tagliatalata-Scafati, Orazio. (2008). *Modern alkaloids : structure, isolation, synthesis and biology*. Wiley-VCH.
- Fenton, S. E., Ducatman, A., Boobis, A., DeWitt, J. C., Lau, C., Ng, C., Smith, J. S., & Roberts, S. M. (2021). Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. In *Environmental Toxicology and Chemistry* (Vol. 40, Issue 3, pp. 606–630). Wiley Blackwell. <https://doi.org/10.1002/etc.4890>
- Ferrari, F., Orlando, A., Ricci, Z., & Ronco, C. (2019). Persistent pollutants: Focus on perfluorinated compounds and kidney. In *Current Opinion in Critical Care* (Vol. 25, Issue 6, pp. 539–549). Lippincott Williams and Wilkins. <https://doi.org/10.1097/MCC.0000000000000658>
- Glüge, J., Scheringer, M., Cousins, I. T., Dewitt, J. C., Goldenman, G., Herzke, D., Lohmann, R., Ng, C. A., Trier, X., & Wang, Z. (2020). An overview of the uses of per- And polyfluoroalkyl substances (PFAS). *Environmental Science: Processes and Impacts*, 22(12), 2345–2373. <https://doi.org/10.1039/d0em00291g>
- Halder, A. K., & Cordeiro, M. N. D. S. (2019). Development of multi-target chemometric models for the inhibition of class I PI3K enzyme isoforms: A case study using QSAR-Co tool. *International Journal of Molecular Sciences*, 20(17). <https://doi.org/10.3390/ijms20174191>
- Halder, A. K., & Dias Soeiro Cordeiro, M. N. (2021). QSAR-Co-X: an open source toolkit for multitarget QSAR modelling. *Journal of Cheminformatics*, 13(1), 1–18. <https://doi.org/10.1186/s13321-021-00508-0>
- Hamadache, M., Amrane, A., Othmane, B., Hanini, S., Khaouane, L., & Si-Moussa, C. (2017). *Environmental Toxicity of Pesticides, and Its Modeling by QSAR Approaches* (pp. 471–501). [https://doi.org/10.1007/978-3-319-56850-8\\_13](https://doi.org/10.1007/978-3-319-56850-8_13)
- Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, 26(6), 822–830. <https://doi.org/10.1093/bioinformatics/btq037>
- Houck, K. A., Patlewicz, G., Richard, A. M., Williams, A. J., Shobair, M. A., Smeltz, M., Clifton, M. S., Wetmore, B., Medvedev, A., & Makarov, S. (2021). Bioactivity profiling of per- and polyfluoroalkyl substances (PFAS) identifies potential toxicity pathways related to molecular structure. *Toxicology*, 457(April), 152789. <https://doi.org/10.1016/j.tox.2021.152789>

- Jain, A. N., & Nicholls, A. (2008). Recommendations for evaluation of computational methods. *Journal of Computer-Aided Molecular Design*, 22(3–4), 133–139. <https://doi.org/10.1007/s10822-008-9196-5>
- Kirk, A. B., Michelsen-Correa, S., Rosen, C., Martin, C. F., & Blumberg, B. (2021). PFAS and Potential Adverse Effects on Bone and Adipose Tissue through Interactions with PPAR $\gamma$ . *Endocrinology (United States)*, 162(12). <https://doi.org/10.1210/endocr/bqab194>
- Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Research*, 44(D1), D1075–D1079. <https://doi.org/10.1093/nar/gkv1075>
- Li, W., Hu, Y., & Bischel, H. N. (2021). In-vitro and in-silico assessment of per- and polyfluoroalkyl substances (PFAS) in aqueous film-forming foam (AFFF) binding to human serum albumin. *Toxics*, 9(3). <https://doi.org/10.3390/toxics9030063>
- Liu, X., Fang, M., Xu, F., & Chen, D. (2019). Trends in Analytical Chemistry Characterization of the binding of per- and poly- fluorinated substances to proteins : A methodological review. *Trends in Analytical Chemistry*, 116, 177–185. <https://doi.org/10.1016/j.trac.2019.05.017>
- López-López, E., Naveja, J. J., & Medina-Franco, J. L. (2019). DataWarrior: an evaluation of the open-source drug discovery tool. *Expert Opinion on Drug Discovery*, 14(4), 335–341. <https://doi.org/10.1080/17460441.2019.1581170>
- Macari, G., Toti, D., Pasquadibisceglie, A., & Polticelli, F. (2020). DockingApp RF: A State-of-the-Art Novel Scoring Function for Molecular Docking in a User-Friendly Interface to AutoDock Vina. *International Journal of Molecular Sciences*, 21(24), 1–17. <https://doi.org/10.3390/ijms21249548>
- Mezey, P. G. (2008). Chimie informatique Molecular symmetry deficiency and shape deviation measures. *Chimie Informatique*, 320–321.
- Ng, C. A., & Hungerbuehler, K. (2015). Exploring the Use of Molecular Docking to Identify Bioaccumulative Perfluorinated Alkyl Acids (PFAAs). *Environmental Science and Technology*, 49(20), 12306–12314. <https://doi.org/10.1021/acs.est.5b03000>
- NTP. (2006). *ICCVAM TEST METHOD EVALUATION REPORT IN VITRO CYTOTOXICITY TEST METHODS FOR ESTIMATING STARTING DOSES FOR ACUTE ORAL SYSTEMIC TOXICITY TESTING*. <http://iccvam.niehs.nih.gov/methods/invitro.htm>
- OECD. (2012). *The OECD QSAR Toolbox for Grouping Chemicals into Categories*. [https://community.oecd.org/community/toolbox\\_forum](https://community.oecd.org/community/toolbox_forum)
- Olsen, G. W., Burris, J. M., Ehresman, D. J., Froehlich, J. W., Seacat, A. M., Butenhoff, J. L., & Zobel, L. R. (2007). *Half-Life of Serum Elimination of Perfluorooctanesulfonate , Perfluorohexanesulfonate , and Perfluorooctanoate in Retired Fluorochemical Production Workers*. 9, 1298–1305. <https://doi.org/10.1289/ehp.10009>
- Otagiri, M., & Giam Chuang, V. T. (2016). Albumin in medicine: Pathological and clinical applications. In *Albumin in Medicine: Pathological and Clinical Applications*. Springer Singapore. <https://doi.org/10.1007/978-981-10-2116-9>
- Pelch, K. E., Reade, A., Wolffe, T. A. M., & Kwiatkowski, C. F. (2019). PFAS health effects database: Protocol for a systematic evidence map. In *Environment International* (Vol. 130). Elsevier Ltd. <https://doi.org/10.1016/j.envint.2019.05.045>
- Salahuddin, Shaharyar, M., & Mazumder, A. (2017). Benzimidazoles: A biologically active compounds. In *Arabian Journal of Chemistry* (Vol. 10, pp. S157–S173). Elsevier B.V. <https://doi.org/10.1016/j.arabjc.2012.07.017>

- Salvalaglio, M., Muscionico, I., & Cavallotti, C. (2010). Determination of energies and sites of binding of PFOA and PFOS to human serum albumin. *Journal of Physical Chemistry B*, 114(46), 14860–14874. <https://doi.org/10.1021/jp106584b>
- Sanders, E. R. (2012). Aseptic laboratory techniques: plating methods. *Journal of Visualized Experiments : JoVE*, 63, e3064. <https://doi.org/10.3791/3064>
- Schymanski, E. L., Chirsir, P., Kondic, T., Thiessen, P. A., Zhang, J., & Bolton, E. E. (2022). *PFAS and Fluorinated Compounds in PubChem Tree*.
- Simard, J. R., Zunszain, P. A., Hamilton, J. A., & Curry, S. (2006). Location of High and Low Affinity Fatty Acid Binding Sites on Human Serum Albumin Revealed by NMR Drug-competition Analysis. *Journal of Molecular Biology*, 361(2), 336–351. <https://doi.org/10.1016/j.jmb.2006.06.028>
- Singam, A., Ramaprasad, E., Tachachartvanich, P., Fourches, D., Soshilov, A., Hsieh, J. C. Y., la Merrill, M., Smith, M. T., & Durkin, K. A. (2020). Structure-Based Virtual Screening of Perfluoroalkyl and Polyfluoroalkyl Substances (PFASs) as Endocrine Disruptors of Androgen Receptor Activity Using Molecular Docking and Machine Learning. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.11886702.v1>
- Søderstrøm, S., Lille-Langøy, R., Yadetie, F., Rauch, M., Milinski, A., Dejaegere, A., Stote, R. H., Goksøyr, A., & Karlsen, O. A. (2022). Agonistic and potentiating effects of perfluoroalkyl substances (PFAS) on the Atlantic cod (*Gadus morhua*) peroxisome proliferator-activated receptors (Ppars). *Environment International*, 163. <https://doi.org/10.1016/j.envint.2022.107203>
- Tice, R. R., Austin, C. P., Kavlock, R. J., & Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: A Tox21 update. In *Environmental Health Perspectives* (Vol. 121, Issue 7, pp. 756–765). <https://doi.org/10.1289/ehp.1205784>
- Tyagi, S., Gupta, P., Saini, A., Kaushal, C., & Sharma, S. (2011). The peroxisome proliferator-activated receptor: A family of nuclear receptors role in various diseases. *Journal of Advanced Pharmaceutical Technology and Research*, 2(4), 236–240. <https://doi.org/10.4103/2231-4040.90879>
- vanden Heuvel, J. P., Thompson, J. T., Frame, S. R. S. R., & Gillies, P. J. (2006). Differential activation of nuclear receptors by perfluorinated fatty acid analogs and natural fatty acids: A comparison of human, mouse, and rat peroxisome proliferator-activated receptor- $\alpha$ , - $\beta$ , and - $\gamma$ , liver X receptor- $\beta$ , and retinoid X receptor- $\alpha$ . *Toxicological Sciences*, 92(2), 476–489. <https://doi.org/10.1093/toxsci/kfl014>
- Velkov, T. (2013). Interactions between human liver fatty acid binding protein and peroxisome proliferator activated receptor selective drugs. *PPAR Research*. <https://doi.org/10.1155/2013/938401>
- Wang, Z., Buser, A. M., Cousins, I. T., Demattio, S., Drost, W., Johansson, O., Ohno, K., Patlewicz, G., Richard, A. M., Walker, G. W., White, G. S., & Leinala, E. (2021). A New OECD Definition for Per- And Polyfluoroalkyl Substances. *Environmental Science and Technology*, 55(23), 15575–15578. <https://doi.org/10.1021/acs.est.1c06896>
- Wang, Z., Dewitt, J. C., Higgins, C. P., & Cousins, I. T. (2017). A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environmental Science and Technology*, 51(5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>
- Wilks, S. S. (1932). CERTAIN GENERALIZATIONS IN THE ANALYSIS OF VARIANCE\*. *Biometrika*, 24, 471–494. <https://academic.oup.com/biomet/article/24/3-4/471/378918>

- Willemsen, J. A. R., & Bourg, I. C. (2021). Molecular dynamics simulation of the adsorption of per- and polyfluoroalkyl substances (PFASs) on smectite clay. *Journal of Colloid and Interface Science*, 585, 337–346. <https://doi.org/10.1016/j.jcis.2020.11.071>
- Wolfrum, C., Borrmann, C. M., Bö Rchers ‡, T., & Spener, F. (2001). Fatty acids and hypolipidemic drugs regulate peroxisome proliferator-activated receptors-and-mediated gene expression via liver fatty acid binding protein: A signaling path to the nucleus. In *PNAS* (Vol. 98, Issue 5). [www.pnas.org/doi/10.1073/pnas.051619898](http://www.pnas.org/doi/10.1073/pnas.051619898)
- Woodcroft, M. W., Ellis, D. A., Rafferty, S. P., Burns, D. C., March, R. E., Stock, N. L., Trumpour, K. S., Yee, J., & Munro, K. (2010). Experimental characterization of the mechanism of perfluorocarboxylic acids' liver protein bioaccumulation: The key role of the neutral species. *Environmental Toxicology and Chemistry*, 29(8), 1669–1677. <https://doi.org/10.1002/etc.199>
- Yang, D., Han, J., Hall, D. R., Sun, J., Fu, J., Kutarna, S., Houck, K. A., Lalone, C. A., Doering, J. A., Ng, C. A., & Peng, H. (2020). *Nontarget Screening of Per- and Poly fl uoroalkyl Substances Binding to Human Liver Fatty Acid Binding Protein*. <https://doi.org/10.1021/acs.est.0c00049>
- Zhang, L., Ren, X. M., & Guo, L. H. (2013). Structure-based investigation on the interaction of perfluorinated compounds with human liver fatty acid binding protein. *Environmental Science and Technology*, 47(19), 11293–11301. <https://doi.org/10.1021/es4026722>

## CHAPTER 5: CONCLUDING REMARK

More chemicals are developed, used, and discharged into the environment to better serve modern life. Under the current legal framework, chemical of emerging concern may continue to be manufactured and pose health risks to human and/or ecological species. The burden of chemical management, disease treatment, and decontamination falls on the downstream users and impacted communities that are located close to the pollution sites. To understand environmental and public health impacts of chemicals that are produced and to prevent regretful production and dependencies on bioaccumulative and toxic chemicals, I present this dissertation with the following concluding remarks.

I develop and present a more efficient and comprehensive safety assessment pipeline for TrOCs in nutrient recovery insects using HRMS and suspect-screening approach. In comparison to targeted analysis, this pipeline aims to reduce bias, which does not pre-select chemical targets of interest. In Chapter 2, I used this approach to identify and semi-quantitatively evaluate the presence of TrOCs that are bioaccumulative, persistent, and toxic in an agricultural waste (e.g., almond hulls). Among all 46 TrOCs suspects detected in the hulls, only bifenthrin bioaccumulated in BSFL tissue. To further expand chemical safety assessment for unknown TrOCs, I performed nontarget analysis and found an emerging class of PFAS (hydrogen-substituted polyfluoroalkyl carboxylic acids; H-PFCAs) sourced from polymer degradation bioaccumulated in BSFL. Moving forward, I recommend using this chemical safety assessments pipeline when assessing novel feed and food products that pose unknown health risks. Future work could focus on establishing a TrOCs chemical suspect library following the FAIR (Findability, Accessibility, Interoperability, Reusability) principal. Ultimately, improved chemical risk assessment workflows may provide information before hazardous chemical production infrastructure is established

I evaluate bioaccumulation potential and mechanisms of PFAS mixture in commercial product using model protein. As a rapid expanding sub-class contaminant of TrOCs, toxicity and bioaccumulation potential of most PFAS are unknown. In Chapter 3, I used suspect-screening and nontarget analysis in addition to targeted approach to include the analysis for less known and novel PFAS. Firstly, I identified at least five PFAS, including two short-chain ones (with less than five perfluorocarbons) ultrastrong bound to HSA. Ultrastrong bound PFAS were released with an additional acid hydrolysis step after the extractions of noncovalent bound PFAS. This observation raises concerns about PFAS toxicity pathways via ultrastrong or even potentially covalent binding. Secondly, I explored and validated a computational tool, molecular docking, for bioaccumulation potential prediction. Molecular docking predicted HSA binding scores were compared with experimental data, which can be used to identify if previously uncharacterized PFAS are likely to bind with HSA. Altogether, this Chapter offers a highthroughput workflow for evaluating bioaccumulation potentials of thousands of TrOCs like PFAS in comparable biological tissues.

I develop and evaluate QSAR models for predicting PFAS toxicity/bioactivity including protein docking scores as a part of the structural descriptors. With advancements in analytical techniques, increasing numbers and diversity of PFAS structures are detected in the environment. Investigating health impacts of TrOCs like PFAS one-by-one is not feasible. In Chapter 4, I developed a PFAS toxicity quantitative structure activity relationship (QSAR) model and identify the importance of protein bindings to toxicity predictions. Using 796 bioactivity responses and descriptive inputs with machine learning models (i.e., GB and RF), I predicted bioactivity/toxicity of 2,649 PFAS with excellent performance. Using linear model, I identified the importance of HSA binding and PPAR  $\gamma$  to PFAS toxicity/bioactivity. Moreover, PFAS produced by the manufacturers are structurally more diverse than the ones being tested for bioactivity.

Future research should focus on data collection and QSAR model development based on structural similarity maps.

## APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

### Table of content

- A1 Detailed chemicals used for quantification
- A2 Lipid quantification in BSFL
- A3 BSFL depuration method

### Figures

- Figure A1. Diagram of BSFL depuration process
- Figure A2. Feed mass loss and larvae growth
- Figure A3. Spike recovery test of pesticides and PFAS in almond hulls and BSFL
- Figure A4. Relative responses of all chemicals qualified in almond hulls
- Figure A5. PFAS concentration in pre-spiked hull #5 and BSFL
- Figure A6. KMD plots of Hull #5 as received and 24 PFAS analytical standards
- Figure A7. Qualification of hydrogen substituted PFCA in almond hull and BSFL
- Figure A8. Annotated MS/MS spectrum of hydrogen substituted PFCA in standards

### Tables

- Table A1.1 BSFL growth information for Stage 2 experiments
- Table A1.2 BSFL growth information for Stage 3 experiments
- Table A2. LC instrumental parameter
- Table A3. GC-EI instrumental parameter
- Table A4. GC-NCI instrumental parameter
- Table A5. Suspect screening parameter setup
- Table A6. County pyrethroid pesticide usage report in 2017
- Table A7. Information of the residual chemicals qualified in six almond hull types in Stage 1 experiments

## **A1. Detailed chemicals used for quantification**

The following analytical standards were purchased from Supelco Inc. (Bellefonte, PA) and used to quantify: bifenthrin (34314); bioallethrin (31489); chlorothalonil (36791); chlorpyrifos (45395); cyfluthrin (46003); cyhalothrin (74767); cypermethrin (36128); cyphenothrin (46037); esfenvalerate (46277). Quantification of residual pesticides was performed with 9-point calibration curve (0.1-500 ng/mL). Whole limits of quantification (LOQ) were determined based on ten times of signal-to-noise ratio (S/N) in BSFL tissue samples, which ranged from 0.5 to 4 ng/g-dw. The LOQ for the same compounds in almond hulls were at least half of the level in BSFL tissues. The LOQ of tetramethrin and cypermethrin were 4 ng/g-dw. The LOQ of cyphenothrin and permethrin were 1.5 ng/g-dw. The LOQ of the rest of pesticides quantified were 1.5 ng/g-dw.

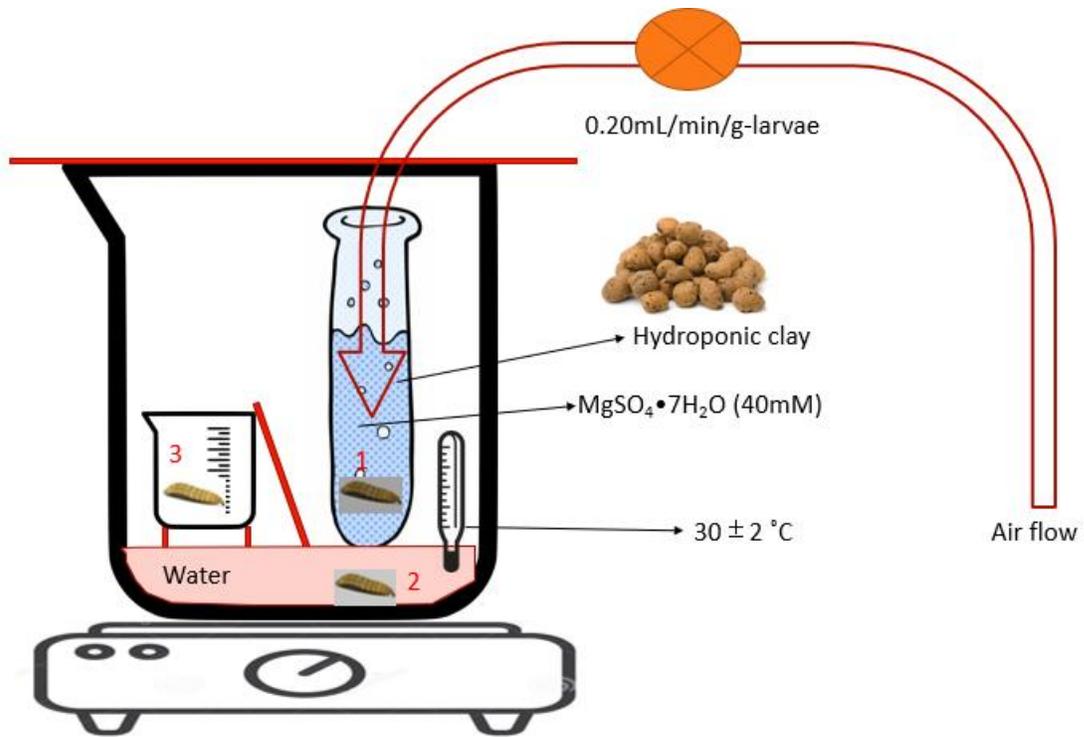
The following PFAS spiked in almond hulls for bioaccumulation study were purchased from Synquest Laboratories Inc. (Alachua, FL): ammonium perfluoro(2-methyl-3-oxahexanoate) (GenX; 2122-3-09); heptafluorobutyric acid (PFBA; 2121-3-34); perfluoro-n-octanoic acid (PFOA; 2121-3-18); perfluorooctanesulfonic acid (PFOS; 6164-3-08). Similarly, quantification of PFAS was performed with 10-point calibration curve (0.5-250 ng/mL). The LOQ were determined ranging from 0.1 to 1 ng/g-dw in BSFL tissues. The LOQ of PFBA and GenX were 1 and 2 ng/g-dw respectively, and the LOQ of the rest of PFAS were 0.5 ng/g-dw. We also purchased 8H-perfluorooctanoic acid (8H-PFCA; 2121-392) and 9H-perfluorononanoic acid (9H-PFCA; 2121-3-23) from Synquest Laboratories for structural confirmations.

## **A2. Lipid quantification in BSFL**

Lipid extraction was performed on the larvae harvested from the Stage 2 experiment and the commercially available larvae-Tasty Grubs™ Larvae (Tasty Worms, GA, USA), using the method by Folch et al. (1957). Specifically, we added 20 mL of chloroform: methanol (2:1, v:v) into 1 g of dry ground larvae and homogenized for an hour. Then we added 4 mL of distilled water to create the tri-phasic system. Impurities, such as salt, protein, chitosan and so on are moved to the aqueous and interphase layer. After centrifugation for 10 min (4,000 rpm), the top layer of lipid in chloroform was displaced in pre-weighed weighing glasses. The weighing glasses was placed in the chemical hood for 24 hr to evaporate chloroform. Finally, the lipid in larvae was determined gravimetrically.

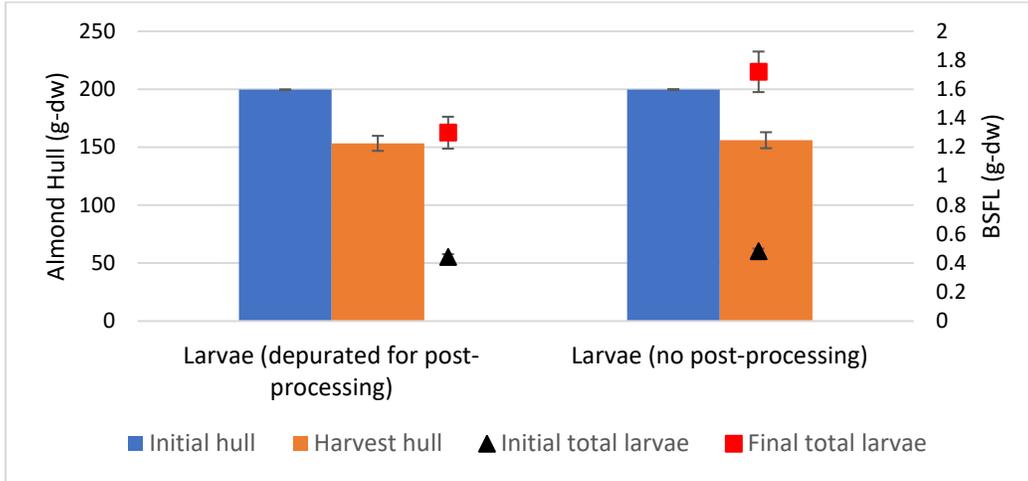
### **A3. BSFL depuration method**

We created BSFL depuration setups using 250 mL beakers shown in the diagram below. The whole depuration process took 24 hr. First, we placed harvested BSFL in a 50 mL centrifuge tubes filled with 25 mL of 40 mM of magnesium sulfate heptahydrate (230391; Sigma-Aldrich, MO, USA). We used the salt solution here to promote osmotic pressure induced depuration (Greenfield et al., 2014; Portalatin & Winstead, 2012). We also placed 10±1 g of hydroponic clay as ladders for larvae to exit and to create additional air pockets as ~1 mL/min of filtered air bubble into the salt solution, labeled as 1 in the diagram. The whole chamber was kept at 30±2 °C. Second, as the larvae moved out of the centrifuge tubes, it was rinsed in deionized water in the beaker, labeled as 2 in the diagram. Third, to avoid excessively warm water larvae climbed up the ramp to a dry 50 mL beaker, which was our collection beaker for the depurated larvae, labeled as 3 in the diagram. We compared this method with blanching at 60 °C (30 minutes) and 100 °C (60 seconds). The blanching methods yielded less than 10% body mass decreasing, and dark gut content was also visually apparent in the larvae collected from the blanching methods. We did not depurate BSFL directly by leaving the larvae in a still container as cannibalism and feeding on gut content were observed.

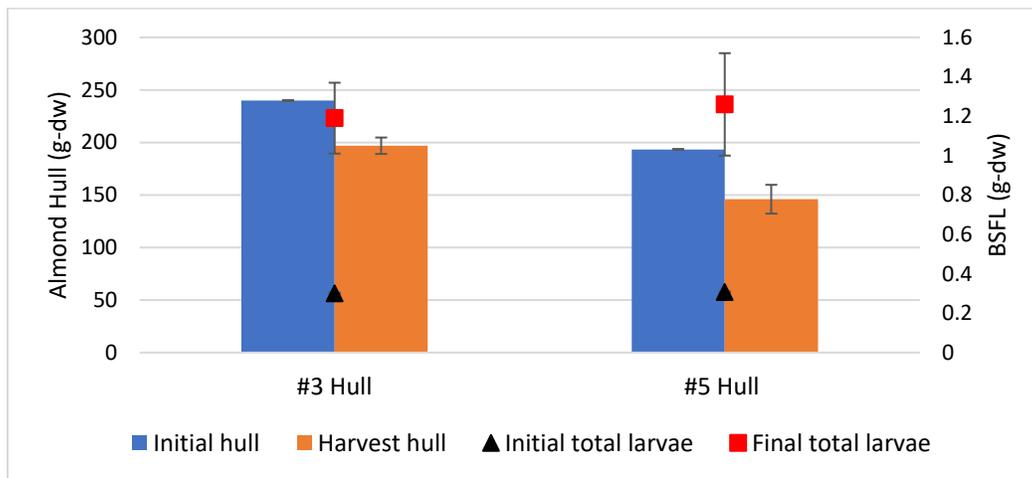


**Figure A1.** Diagram for black soldier fly larvae (BSFL) depuration process

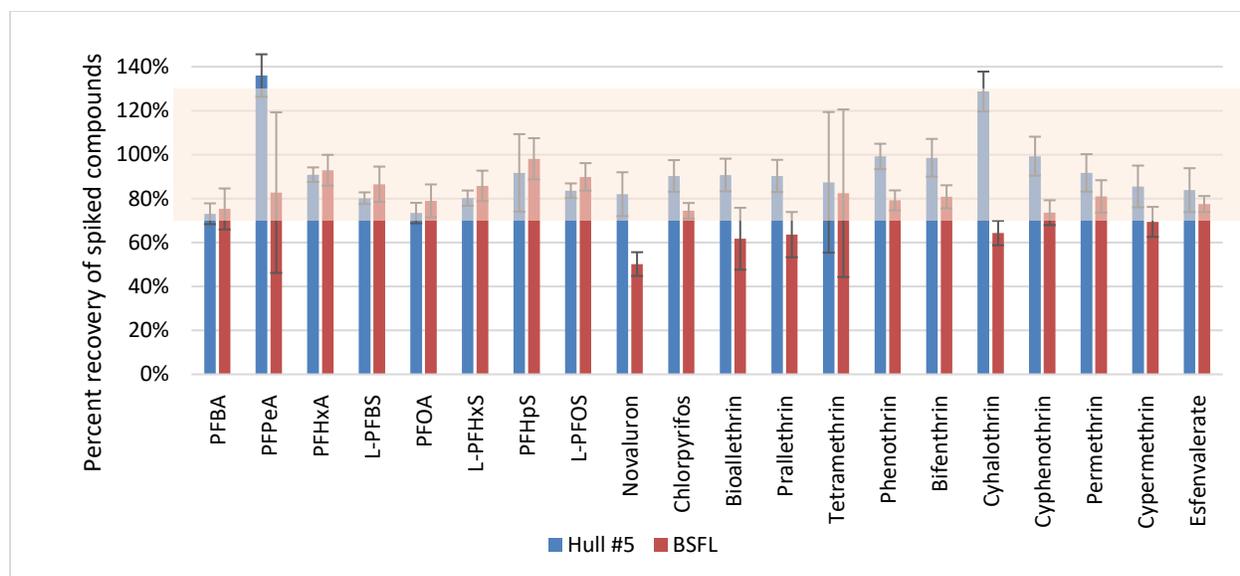
(a)



(b)

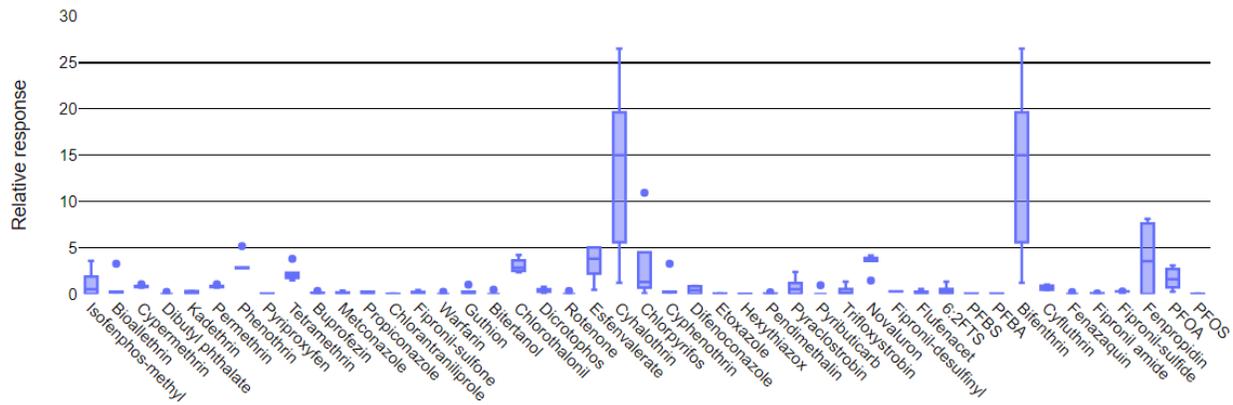


**Figure A2.** Dry mass changes of hulls and BSFL during the bioaccumulation study period (14-day) in (a) Stage 2 experiments, where Hulls #3 and #5 were used to produce BSFL and (b) Stage 3 experiments, where only Hull #5 was used to produce BSFL, and we depurated BSFL from four reactors to assess larval tissue.



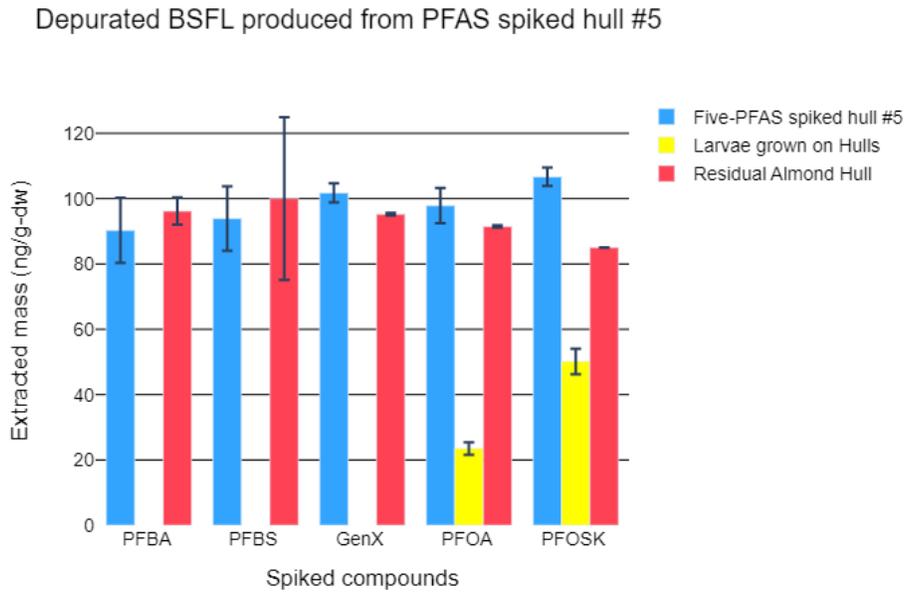
**Figure A3.** Spike recovery test (200 ng) of pesticides and PFAS in Hull #5 and BSFL (Tasty Grubs™ Larvae). The percent recovery was calculated by taking the concentration differences between pre-spiked and non-spiked sample matrices over the concentration differences between post-spiked and non-spiked sample matrices. The shaded area represents the conventional acceptable recovery range (70-130%) in analytical chemistry.

Residual chemicals detected in six almond hulls

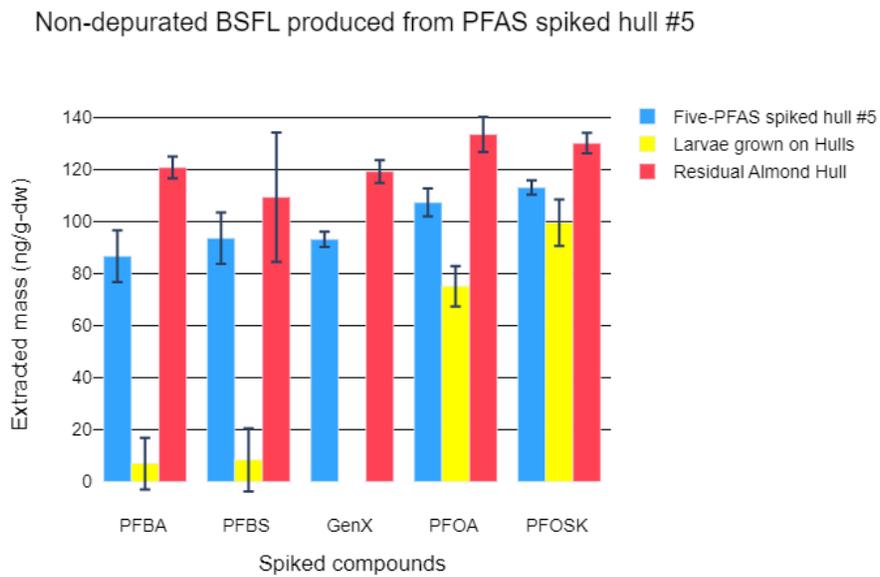


**Figure A4.** Box-plot (5th, 25th, 50th, 75th, 95th percentiles) showing relative responses of all chemicals qualified in six almond hull types. The relative responses were calculated by dividing the peak area of each compound over their corresponding ISTD.

(a)

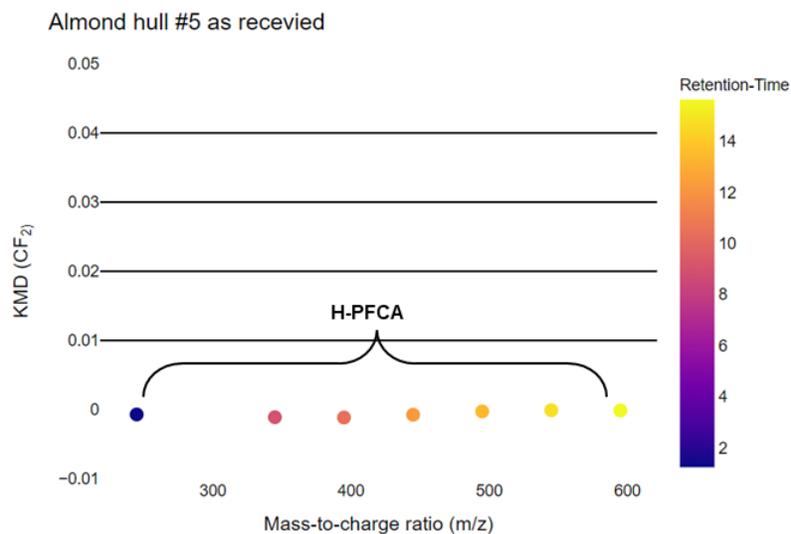


(b)

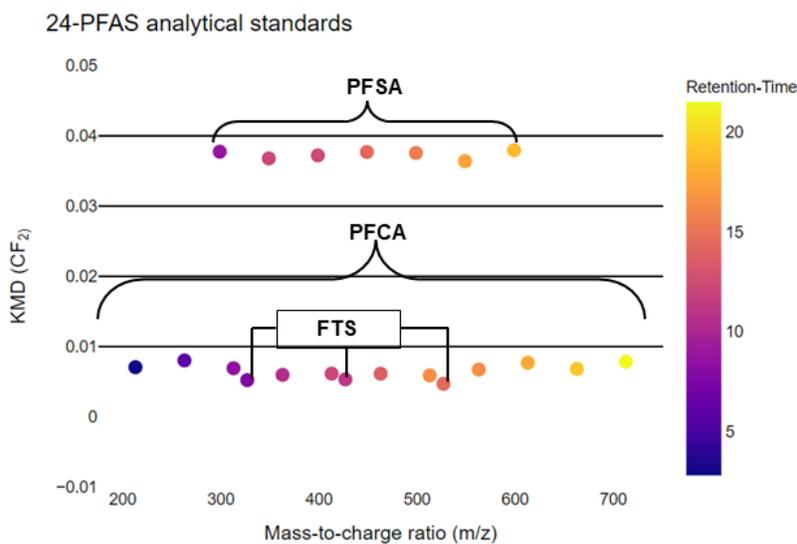


**Figure A5.** Concentrations of five targeted PFAS in hull #5 spiked with each PFAS, residual hulls following digestion by BSFL, and BSFL which were (a) not de-purated; and (b) de-purated.

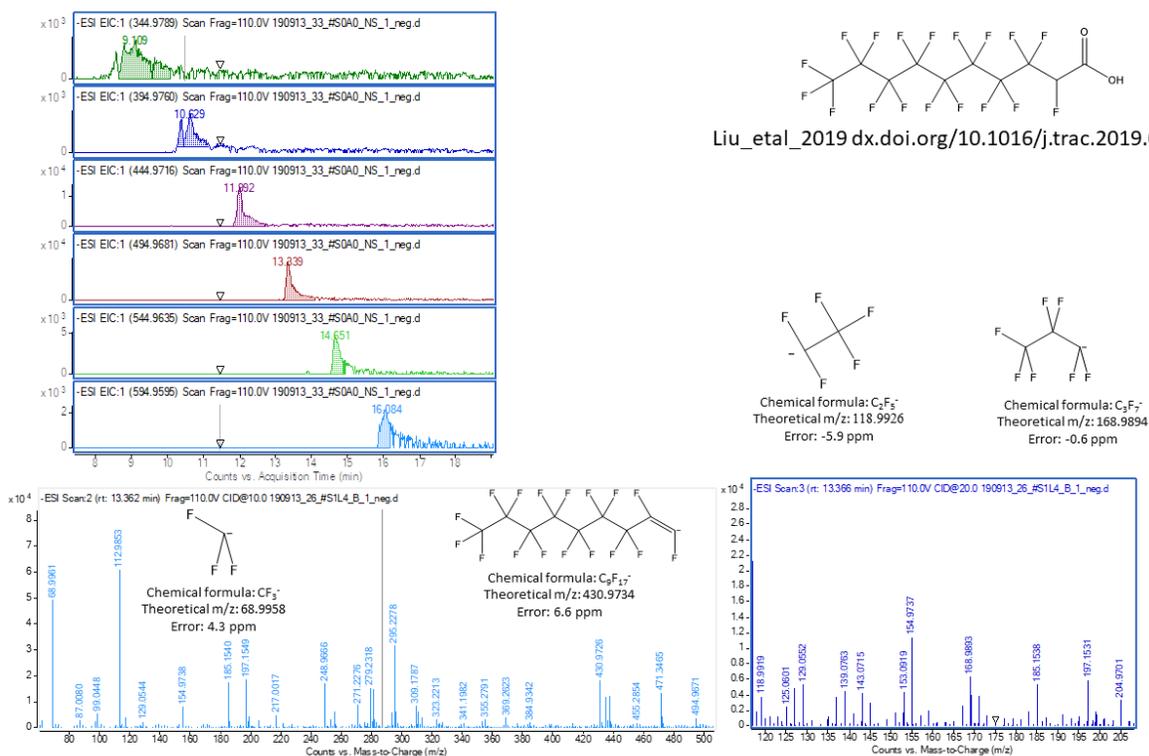
(a)



(b)

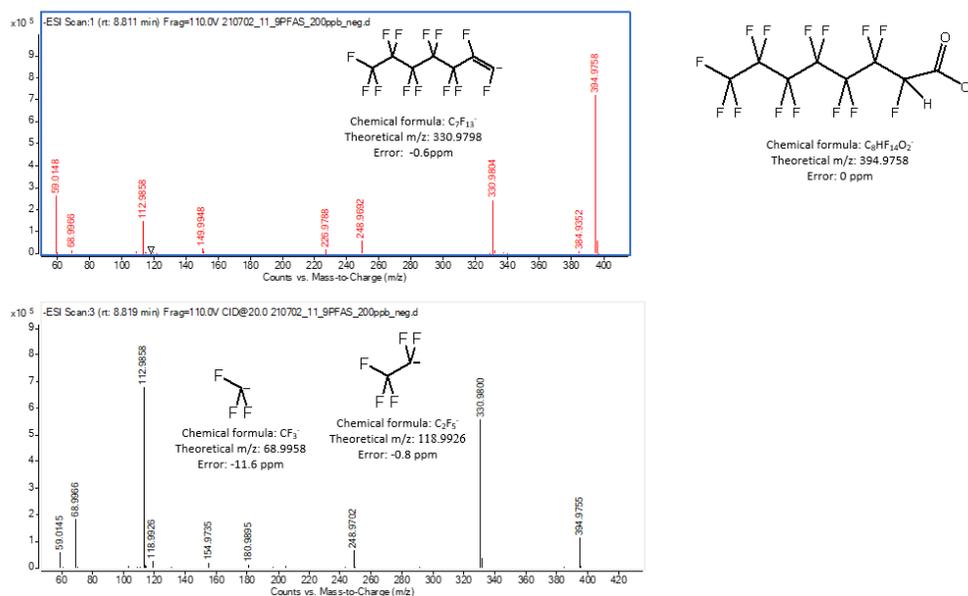


**Figure A6.** Kendric mass defect (KMD) analysis plots of (a) Hull #5 as received and (b) the 24-PFAS analytical standards mix used for spike-recovery tests. The color of each dot represents its retention time. No PFAS homologue series were detected in the five individual PFAS standards used to spike hulls in Stage 3 bioaccumulation experiments.

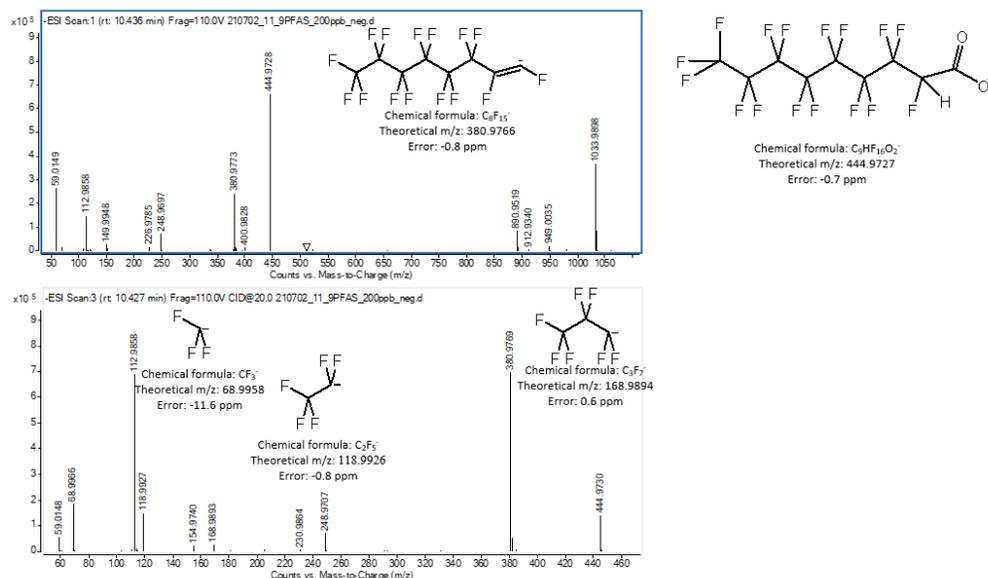


**Figure A7.** Qualification of a hydrogen-substituted PFCA (H-PFCA) to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The top left inset shows extracted chromatograms for the molecular ions of six H-PFCAs identified in the almond hulls using KMD analysis. The bottom insets are examples of MS/MS of a BSFL extract and fragment ions of HPFDA manually assigned at 10 eV (lower left) and 20 eV (lower right) collision energy acquisition at the retention time (rt) shown.

(a)



(b)



**Figure A8.** Mass spectra of (a) 8H-perfluorooctanoic acid (HPFOA; or 8H-PFCA) and (b) 9H-perfluorononanoic acid (HPFNA; or 9H-PFCA) standards, with manually assigned structures.

Using theoretical molecular mass, the top spectra were extracted at 0 eV collision energy and the bottom spectra were extracted at 20 eV collision energy in both (a) and (b).

**Table A1.1** BSFL growth information for Stage 2 experiments

Stage 2 experiment (Harvest date 11/6/2018)											
Reactor #	Hull variety	With larvae 1=yes , 0=no	Initial dry weight of hulls (g)	Initial dry weight of larvae (g)	Initial larvae count	Depurated, 1=yes , 0=no	PFAS spiked, 1=yes, 0=no	Hull consumption, dry (g)	Larvae harvested, g DRY	Larvae survival rate (%)	Larvae growth rate (%)
4	Pollinator, Chico 2017	1	239.9±0.1	0.301±0.002	100	1	0	43.0±7.8	1.19±0.8	95.2±4.7%	295±67%
4	Pollinator, Chico 2017	0	234.9±8.5	NA	NA	NA	0	38.5±17.9	NA	NA	NA
4	Monterey, Buttonwillow 2017	1	193.4±0.1	0.307±0.001	100	1	0	47.4±13.7	1.26±0.26	94.3±6.6%	317±21%
4	Monterey, Buttonwillow 2017	0	188.1±6.2	NA	NA	NA	0	35.4±19.0	NA	NA	NA

**Table A1.2** BSFL growth information for Stage 3 experiments

Stage 3 experiment (Harvest date 7/11/2019)												
Reactor #	Hull variety	With larvae, 1=yes, 0=no	Initial dry weight of hulls (g)	Initial dry weight of larvae (g)	Initial larvae count	Depurated 1=yes 0=no	PFAS spiked 1=yes 0=no	Hull consumption, dry	Larvae harvested, g DRY	Larvae survival rate (%)	Larvae growth rate (%)	Gut content % (dw/dw)
4	Monterey, Buttonwillow 2017	1.00	199.6±0.2	0.442±0.078	100	1	1	46.3±6.5	1.30±0.11	93.7±1.5%	275±23%	21.0±2.2%
4	Monterey, Buttonwillow 2017	1.00	199.7±0.2	0.481±0.051	100	0	1	47.3±6.9	1.72±0.14	95.3±2.9%	258±18%	NA
1	Monterey, Buttonwillow 2017	0.00	196.4	NA	NA	NA	1	18.1	NA	NA	NA	NA
1	Monterey, Buttonwillow 2017	0.00	205.8	NA	NA	NA	0	29.6	NA	NA	NA	NA

**Table A2.** LC instrumental parameter

<b>LC-QTOF-MS Method</b>	
Injection Volume	10 $\mu$ L
<b>LC Settings</b>	
Mobile Phases	A: H <sub>2</sub> O + 20 mM Ammonium Acetate
	B: Acetonitrile
Solvent Flow	0.40 mL/min
Gradient for samples	25% B for 1.5 min
	25%-90% B in 25 min
	100% B for 5 min
	equilibration to initial conditions for 3 min
Column	ZORBAX RRHD Eclipse Plus 95Å, 2.1 x 150 mm, 1.8 $\mu$ m)
Column Temperature	30°C
<b>MS Settings</b>	
Gas Temperature	300 °C
Drying Gas Flow	12 l/min
Nebulizer	25 psig
Sheath Gas Temperature	350 °C
Sheath Gas Flow	11 l/min
Vcap	3000 (neg)
Fragmentor	110 V
scan range	50-1200 m/z
scan speed	4 spectra/s
All-Ions Acquisition	Collision Energy (CE): 0, 10, 20, 40
Reference Mass Correction	neg masses: 112.9855, 1033.9881
	pos masses: 121.0509, 922.0098

**Table A3.** GC-EI instrumental parameter

<b>Table S2.2 GC (EI) acquisition parameter</b>	
Run Time	78.5 min
Post Run Time	0 min
Oven temperature setpoint	On
(Initial)	35 °C
Hold Time	3 min
Post Run	50 °C
Program	
#1 Rate	4 °C/min
#1 Value	325 °C
#1 Hold Time	3 min
Equilibration Time	3 min
Max Temperature	325 °C

**Table A4.** GC-NCI instrumental parameter

<b>Table S2.3 GC (NCI) Acquisition parameter</b>	
Run Time	36.351 min
Post Run Time	0 min
Oven temperature setpoint	On
(Initial)	100 °C
Hold Time	1 min
Post Run	50 °C
Program	
#1 Rate	15 °C/min
#1 Value	200 °C
#1 Hold Time	0 min
#2 Rate	3.8 °C/min
#2 Value	290 °C
#2 Hold Time	0 min
#3 Rate	10 °C/min
#3 Value	300 °C
#3 Hold Time	4 min
Equilibration	3 min
Max Temperature	325 °C

**Table A5.** Suspect screening parameter setup

Software	Agilent MassHunter Qualitative Analysis (B.08.00)
Workflow	Find Compounds by Formula
Values to Match	Accurate mass
Libraries	In-house PFASs database (3794 PFASs, 61 compounds with MS/MS spectra from in-house standard, massbank)
	Agilent Pesticide PCDL (852 compounds with MS/MS spectra)
	Water Contaminants PCDL (1083 compounds with MS/MS spectra)
Extraction Algorithm	Agile 2
Match Tolerance Masses	LC: $\pm 10$ ppm; GC: $\pm 5$ ppm
Allowed Adducts (LC)	ESI NEG: [M-H] <sup>-</sup> , [M-CH <sub>3</sub> OOH] <sup>-</sup> ;
	ESI POS: [M+H] <sup>+</sup> , [M+Na] <sup>+</sup> , [M+NH <sub>4</sub> ] <sup>+</sup>
Allowed Charge State	1
Isotope Model	common organic molecules
Peak spacing tolerance	0.0025 m/z, plus 7 ppm
Scoring (Weight)	Mass Score: 100
	Isotope Abundance Score: 50
	Isotope Spacing Score: 60
Find by Formula Score	>80 (out of 100)
Absolute Height	> 1000 counts
Confirm with fragment ions	if MS/MS spectra available
number of most specific ions from MS/MS library	5
RT difference	$\pm 0.2$ min
S/N ratio	not applied
Coelution score	> 85%
Minimum Number of Qualified Fragments	LC: 1 GC: 2

**Table A6.** County pyrethroid pesticide usage report in 2017

	Bifenthrin	Permethrin	Lambda-cyhalothrin	Esfenvalerate
Butte County (lbs)-Hull #3	2402.09	32.57	87.54	296.78
Kern County (lbs)-Hull #5	15420.36	304.07	823.03	2347.47

Note: 2017 Annual Statewide Pesticide Use Report for Almond Orchard in Butte County and Buttonwillow, California (<https://files.cdpr.ca.gov/pub/outgoing/pur/data/>)

**Table A7.** Information of the residual chemicals qualified in six almond hull types in Stage 1 experiments

Compounds	Molecular Formula	Biowin 2 (V 4.10)*	Biowin 3 (V 4.10)*	Log Kow (KOWIN v. 1.68)*	Chronic (or acute) toxicity mg/(kg-day); if LD50 (mg/kg); if LC50 (mg/L)	Test animal	Type	b/vb	p/vp	tox
Bifenthrin	C23H22O2ClF3	0.0031	1.4538	8.2	0.01 (acute)	human	RfD	y	y	y
Phenothrin	C23H26O3	0.9883	2.3167	7.5	50	rat	NOAEL	y	n	n
Permethrin	C21H20O3Cl2	0.5375	1.8801	7.4	25	rat	NOAEL	y	n	n
Etoxazole	C21H23NO2F2	0	2.1224	7.2	4	rat	NOAEL	y	y	n
Cyhalothrin	C23H19NO3ClF3	0.5364	1.3285	6.9	0.001 (chronic)	human	RfD	y	n	y
Esfenvalerate	C25H22NO3Cl	0.9998	2.0115	6.8	0.0018	human	RfD	y	n	y
Trifloxystrobin	C20H19N2O4F3	0.01	1.924	6.6	62.2	rat	NOAEL	y	y	n
Cyphenothrin	C24H25NO3	0.9998	2.179	6.5	318	rat	LD50	y	n	y
Fenpropidin	C19H31N	0.01	2.0531	6.4	0.07 (acute)	human	RfD	y	y	y
Cypermethrin	C22H19NO3Cl2	0.9883	1.7424	6.4	7.5	rat	NOAEL	y	n	n
Kadethrin	C23H24O4S	0.892	2.1982	6.3	650	rat	LD50	y	n	n
Fenazaquin	C20H22N2O	0.4403	2.177	5.8	0.05	human	RfD	y	y	y
Cyfluthrin	C22H18NO3Cl2F	0.0017	1.2957	5.7	0.02 (acute)	human	RfD	y	y	y
Hexythiazox	C17H21N2O2ClS	0.02	2.2127	5.6	23	rat	NOAEL	y	y	n
Pyriproxyfen	C20H19NO3	1	2.1224	5.6	35.1	rat	NOAEL	y	n	n
Tetramethrin	C19H25NO4	0.659	2.3949	5.5	0.6	mouse	NOAEL	y	n	n
Bioallethrin	C19H26O3	0.6497	2.4365	5.5	5.9	rat	NOAEL	y	n	n
Pyraclostrobin	C19H18N3O4Cl	0.09	2.0774	5.5	3.4	rat	NOAEL	y	y	n
Fipronil amide	C12H6N4O2Cl2F6S	0	2.4086	5.4	0.0432	yellow fever mosquito	LC50	y	y	y
Pyributicarb	C18H22N2O2S	0.1434	1.9378	5.3	5000	rat	LD50	y	y	n
Novaluron	C17H9N2O4ClF8	0	2.638	5.3	0.011	human	RfD	y	y	n
Difenoconazole	C19H17N3O3Cl2	0	1.6006	5.2	0.5	human	RfD	y	y	n
Chlorpyrifos	C9H11NO3Cl3PS	1	1.7442	5.1	0.003	human	RfD	y	n	y
Prallethrin	C19H24O3	0.6562	2.4409	4.9	16.3	rat	NOAEL	y	n	n

Fipronil-sulfide	C12H4N4Cl2F6S	0	2.1871	4.8	0.004	yellow fever mosquito	LC50	y	y	y
Pendimethalin	C13H19N3O4	0	1.9537	4.8	250	rat	NOAEL	y	y	n
PFOA	C8HO2F15	0	0.8631	4.8	0.0009	human	NOAEL	y	y	y
Dibutyl phthalate	C16H22O4	1	3.4612	4.6	125	mouse	NOAEL	y	n	n
PFOS	C8HO3F17S	0	0.28887	4.5	0.0006	human	NOAEL	y	y	y
Fipronil-sulfone	C12H4N4O2Cl2F6S	0	2.1409	4.4	184	rat	LD50	n	y	n
Rotenone	C23H22O6	0.9999	1.8648	4.3	0.015 (acute)	human	RfD	n	n	y
Buprofezin	C16H23N3OS	0.22	2.3341	4.3	0.9	rat	NOAEL	n	y	n
Fipronil-desulfinyl	C12H4N4Cl2F6	0	2.2333	4.2	0.032	rat	NOAEL	n	y	y
Metconazole	C17H22N3OCI	0	1.7867	4.2	4.3	rat	NOAEL	n	y	n
Propiconazole	C15H17N3O2Cl2	0	1.8002	4.1	18.1	rat	NOAEL	n	y	n
Bitertanol	C23H19NO3ClF3	0.8401	2.3652	4.1	0.01	human	RfD	n	n	y
Chlorantraniliprole	C18H14N5O2BrCl2	0.01	2.9901	4.0	156	rat	NOAEL	n	y	n
Isofenphos-methyl	C14H22NO4PS	0.9155	2.6071	3.9	0.08	human	NOEL	n	n	n
Chlorothalonil	C8N2Cl4	0.6127	1.6204	3.7	0.015	human	RfD	n	n	y
6:2 FTS	C8H5O3F13S	0	0.872	2.7	0.0039	human	RfD	n	y	y
Guthion	C10H12N3O3PS2	1	2.6517	2.5	0.003	human	MRL	n	n	y
Flufenacet	C14H13N3O2F4S	0	1.364	2.4	0.0017	rat	RfD	n	y	y
Warfarin	C19H16O4	0.9968	2.7426	2.2	0.0003	human	RfD	n	n	y
PFBA	C4HO2F7	0.0002	2.1536	2.1	0.0038	human	rfd	n	y	y
PFBS	C4HO3F9S	0.0052	1.5793	1.8	0.0003	human	rfd	n	y	y
Dicrotophos	C8H16NO5P	1	2.7745	-0.1	0.02	mouse	NOAEL	n	n	y

Note:

\* Values are estimated by EPIWEB4.1 (<https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface>); downloaded and installed on December 11th

## APPENDIX B: SUPPLEMENTARY INFORMATION FOR CHAPTER 3

### Table of content

- B1.1 Details on the serum extractions
- B1.2 Manual annotation of PFAS MS/MS spectra
- B2.1 Materials used in <sup>19</sup>F NMR
- B2.2 Details on <sup>19</sup>F qNMR method
- B2.4 Details on <sup>19</sup>F qNMR data acquisition method
- B3.1 Details on molecular docking
- B3.2 Details on ligand preparation

### Figures

- Figure B1.1 Experimental setup for equilibrium dialysis
- Figure B1.2 Assessment of equilibration time required for equilibrium dialysis
- Figure B1.3 Isotherm models for PFAS adsorbed to HSA
- Figure B1.4 Bound fractions of PFAS compounds in AFFF(df=1E4) to HSA
- Figure B2.1 Crystal structures of Human Serum Albumin (HSA)
- Figure B2.2 Predicted binding scores of 26-PFAS binding to HSA (1E7G)
- Figure B2.3 Molecular docking predicted association constants (K<sub>A</sub>) of 26 PFAS to HSA
- Figure B2.4 Uncorrected K<sub>A</sub> values for 1E7G from this study correlate to the results reported by Ng and Hungerbuehler (2015)
- Figure B3.1 Histogram of molecular docking predicted target PFCA-1E7G binding scores
- Figure B3.2 Histogram of molecular docking predicted target PFSA-1E7G binding scores
- Figure B3.3 Histogram of molecular docking predicted target PFAS(precursor)-1E7G binding scores
- Figure B3.4 Histograms of molecular docking predicted scores for C10-AEnS binding 1E7G
- Figure B3.5 Histograms of molecular docking predicted scores for C12-AEnS binding 1E7G
- Figure B3.6 Histograms of molecular docking predicted scores for Cn-LAS binding 1E7G
- Figure B4.0 Homologous plot of hydrocarbon surfactants in AFFF and HSA aliquots (RT vs m/z)
- Figure B4.(1-19) Annotated MS/MS spectrum of qualified compounds in ESI-

### Tables

- Table B1.1 Mass (ng) of PFAS present in each fraction of equilibrium dialysis extracts
- Table B2.1 Molecular docking searching information for 1E7G
- Table B2.2 Molecular docking searching information for 1A06
- Table B2.3 Molecular docking searching information for 4E99
- Table B5.1 LC-QTOF-MS acquisition method
- Table B5.2 Suspect screening software search algorithm criteria for LC-QTOF-MS acquired data
- Table B5.3 <sup>19</sup>F and <sup>1</sup>H NMR acquisition method
- Table B6.1 List of LC-QTOF-MS targeted compounds, internal standards, LOQ, and extraction recoveries
- Table B6.2 Qualified PFAS in AFFF using suspect-screening
- Table B6.3 Qualified hydrocarbon surfactants in AFFF using suspect-screening
- Table B7.1 AFFF formulation comparison of total fluorine quantified in AFFF using HPLC-QTOF-MS and <sup>19</sup>F NMR
- Table B7.2 Experimentally determined and molecular docking predicted HSA-PFAS association constants

Table B7.3 Kruskal-Wallis one-way pairwise ANOVA analysis on docking predicted HSA (1E7G) binding affinities of long-chain PFAS and hydrocarbon surfactants identified in AFFF (Kruskal-Wallis ANOVA)

### **B1.1 Details on serum extractions**

A schematic of a dialysis cell experiment setup is shown in Figure S1.1. After the system reached equilibrium, 100  $\mu\text{L}$  of solution from the chemical side was transferred along with 25 ng internal standard (ISTD) mix and 100  $\mu\text{L}$  of methanol (MeOH) for HPLC-QTOF-MS analysis. 100  $\mu\text{L}$  of HSA solution was added into an eppendorf tube with 200  $\mu\text{L}$  of 0.1 M formic acid and 1.7 mL cold ( $-20^\circ\text{C}$ ) acetonitrile (ACN) for protein denaturation and precipitation. The sample tube was vortexed and centrifuged ( $16,800 \times g$ ) for 3 min to separate the protein pellet and the aliquot. The aliquot fractions were solvent exchanged into MeOH, concentrated to 200  $\mu\text{L}$ , and equilibrated with 25 ng of ISTD mix. To assess if any ultrastrong or covalent bonds were present, loosely associated PFAS in the HSA pellet were washed with 2 mL of ACN five times and the last fraction was saved to confirm PFAS levels were below detection (vanden Heuvel et al., 1992). After evaporating residual ACN, dry protein pellets were hydrolyzed with 50  $\mu\text{L}$  of 6N HCl in closed cap eppendorf tubes at  $110^\circ\text{C}$  for 24 hr (Hirs et al., 1954; Muñoz et al., 2011; Mustățea et al., 2019; Otter, 2012). Excess HCl was evaporated under  $\text{N}_2$ , and the hydrolyzed pellet was neutralized with NaOH and solvent exchanged to MeOH. After syringe filtration (Captiva Agilent Premium Syringe Filter, 2  $\mu\text{m}$ , regenerated cellulose membrane), protein pellet extracts were equilibrated with internal standards and ready for analysis. Details on the native and internal standards used are given in the Supporting Information S-6.1 LCMS target compounds.

## B1.2 Manual annotation of PFAS MS/MS spectra

For MS/MS spectra annotation, we used an *in-silico* fragmentation tool and literature reported common ESI fragments. We acquired data-dependent MS/MS spectra of suspect-screening qualified compounds using retention time and tentative molecular m/z from the original runs and then used MetFrag Web Tool (<https://msbi.ipb-halle.de/MetFragBeta/>) to identified fragmented ions, following the parameter settings reported by Moschet et al. (2018b). Furthermore, we manually screened the fragment ion list extracted from MS/MS spectra against the “Fragments ESI-” published as supporting information by Barzen-Hanson et al. (2017). In this manual annotation step, we assigned Schymanski's confidence level 3 to the compounds with at least one fragment ion in addition to the presence of molecular ion(s). We assigned the confidence level 2b to the compounds with at least 4 fragmented ions that could piece together the molecular ion. In addition to [M-H], the [M-CH<sub>3</sub>COO] adduct was used as qualifier due to abundant acetate ions from the aqueous mobile phase. Example MS/MS spectra annotations are presented in Figure S-4.1 to S-4.20.

## B2.1 Materials used in $^{19}\text{F}$ NMR

The following materials were used in  $^{19}\text{F}$  NMR analysis of AFFF: (1) solvents: methanol- $d_4$  (MeOD) 99.96%D (Sigma Aldrich: 444758 ); deuterium oxide- $d_2$  99.9%D (Sigma Aldrich: 151882 ); HEPES buffer (Sigma Aldrich: 83264); (2) chemical shift reference compound: fluorotrichloromethane ( $\text{CFCl}_3$ ), 99+%, (Sigma Aldrich: 25499-1); (3) calibrating standard addition compounds: 2,2,2-trifluoroacetamide (TFAcAm), 97%, (Sigma Aldrich: 14465-7); 2,2,2-trifluoroethanol (TFE), 99%, (Polysciences, Inc:612197); (4) target compounds: perfluorooctanoic acid (PFOA), 95%, (Sigma Aldrich: 171468); perfluorooctanesulfonic acid (PFOS), 100  $\mu\text{g}/\text{mL}$ , (Supelco: 33607); perfluoro-2-propoxypropanoic acid (PFPrOPrA, “GenX”), 97%, (Synquest Laboratories: 2121-3-13); 1H,1H,2H,2H-Perfluorodecan-1-ol, 99%, (Synquest Laboratories: 2101-3-95); perfluorobutanesulfonic acid (PFBS), 97%, (Synquest Laboratories: 6164-3-09); perfluoro butanoic acid (PFBA), 98%, (Synquest Laboratories: 2121-3-34); perfluoroglutaric acid, 97%, (Sigma Aldrich: 196908).

## B2.2 <sup>19</sup>F-NMR method

<sup>19</sup>F and <sup>1</sup>H NMR spectroscopy were performed with a 500 MHz Bruker Avance DRX NMR spectrometer equipped with a quad probe with 90° power pulse (acquisition settings shown in Table S-5.3). The total fluorine quantification method was developed based on a report by the United States Naval Research Laboratory with modification (Snow et al., 2017). The chemical shift of the sample was calibrated against fluorotrichloromethane (CFCl<sub>3</sub>) in methanol-d<sub>4</sub> (CD<sub>3</sub>OD) and deuterium oxide-d<sub>2</sub> (D<sub>2</sub>O) at 0 ppm. For total fluorine quantification, trifluoroethanol (TFE) was used as the internal standard in a solvent mixture of CD<sub>3</sub>OD and methanol.

Two internal calibration standards (TFE and TFAcAm) were tested, and they were both calibrated with CFCl<sub>3</sub> (δ<sub>F</sub>= 0 ppm). When no additional PFAS standard was added, TFE/CD<sub>3</sub>OD (δ<sub>F</sub>= -77.08 +/-0.10 ppm) appeared to have more consistent chemical shift compared to TFAcAm/CD<sub>3</sub>OD (δ<sub>F</sub>= -75.78 +/-0.49 ppm). Moreover, as the model compound (e.g., PFOA) was added into the NMR tube, a decrease in intensity for TFAcAm was observed. Therefore, total fluorine quantification was performed solely with TFE. For total fluorine quantification in AFFF, a constant volume of 10% TFE (20 μL) in MeOD was added into NMR quartz tubes (Wilmad: 16-800-338) along with 90 μL of AFFF or calibration standard dilutions, 450 μL MeOD, and 1 mL of CFCl<sub>3</sub> vapor. As shown in Equation (1), the concentration of total fluorine content in AFFF can be determined with fluorine signal ratios of TFE with known fluorine content and AFFF dilutions with varied dilution factors.

$$\frac{C_{TFE(F)} V_{TFE}}{C_{AFFF(F)} V_{AFFF}} = \frac{A_{TFE}}{A_{AFFF}} \quad \text{Equation (1)}$$

In this equation,  $C_{TFE(F)}$  is the fluorine concentration in the TFE dilution (mg/ $\mu$ L),  $V_{TFE}$  is the volume of the TFE dilution ( $\mu$ L),  $C_{AFFF(F)}$  is the fluorine concentration in the AFFF dilution (mg/ $\mu$ L),  $V_{AFFF}$  is the volume of the AFFF dilution ( $\mu$ L),  $A_{TFE}$  is integrated peak area of  $^{19}\text{F}$  resonance caused by TFE, and  $A_{AFFF}$  is the sum of integrated area of  $^{19}\text{F}$  resonance caused by PFAS in AFFF. We tested the accuracy of the method with GenX dilutions, achieving a percent error of  $3.8\pm 3\%$  for triplicated samples ( $R^2 > 0.9995$ ).

### B3.1 Details on molecular docking

We used AutoDock Vina (v1.1.2) to simulated PFAS-protein interactions. AutoDock Vina assumes rigid target proteins and flexible ligands. According to the binding sites of fatty acids, the search area was divided into six parts for seven fatty-acid (FA) binding sites: FA1, FA2, FA3/4, FA5, FA6, and FA7. The 9 best binding modes were selected based on energy minimization (Curry, 2003). The search-box settings for 1E7G, 1AO6, and 4E99 are shown in Table S-2.1, S-2.2, and S-2.3. The docking results are constructed with two types of files: one contains free energy values corresponding to variance of the best binding modes in \*.txt, and the second contains structural conformations in \*.pdbqt corresponding to each binding mode. The structural conformations were visualized in Pymol software. Data cleaning to compile each energy value output into an integrated list was completed using the dataframe package in Python. The equilibrium association constant ( $K_A$ ) was calculated according to Equation (2) where  $\Delta G$  is the Gibb's free energy value, and T is the temperature (in Vina, T=300 K) (Ackley et al., 2004; Ng & Hungerbuehler, 2015).

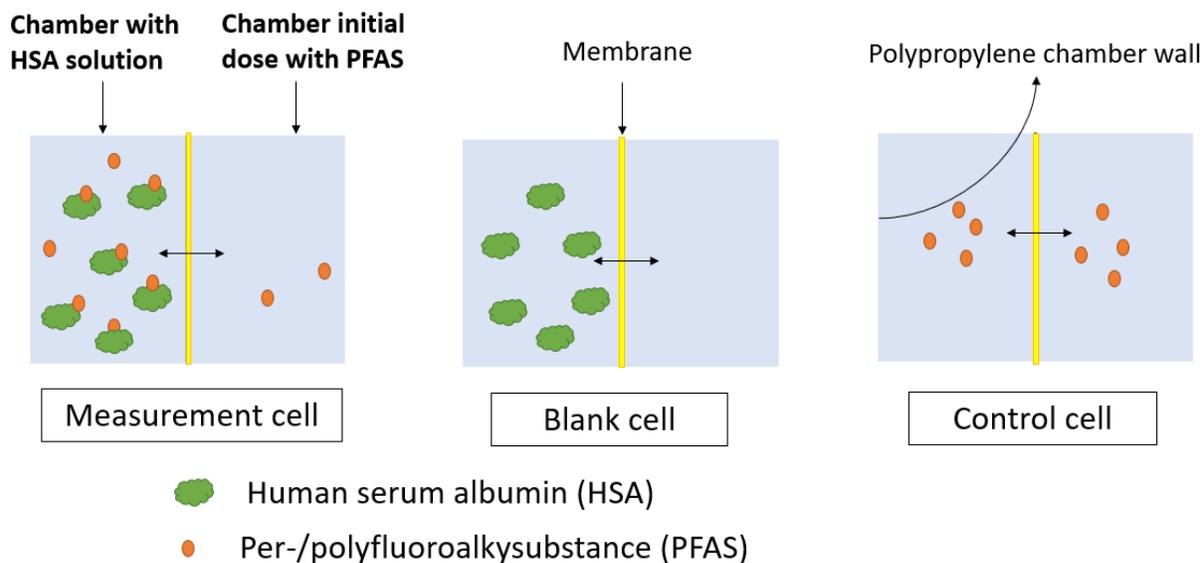
$$K_A = e^{-\frac{\Delta G}{RT}} \quad \text{Equation (2)}$$

The precision and accuracy of the docking simulation was evaluated following the workflow outlined by Ng and Hungerbuehler (2015). The precision of the model was evaluated by repeating the docking simulations of 26 PFAS docking 1E7G at six binding boxes for 100 times. High reproducibility was observed (Figure S-2.2). Accuracy was evaluated by comparing molecular docking predictions to experimentally determined association constants as well as by redocking PFOS to HSA 4E99. 4E99 is a literature-reported HSA crystal structure complexed with PFOS (Luo et al., 2012). First, we artificially removed PFOS and the then assessed the alignment of

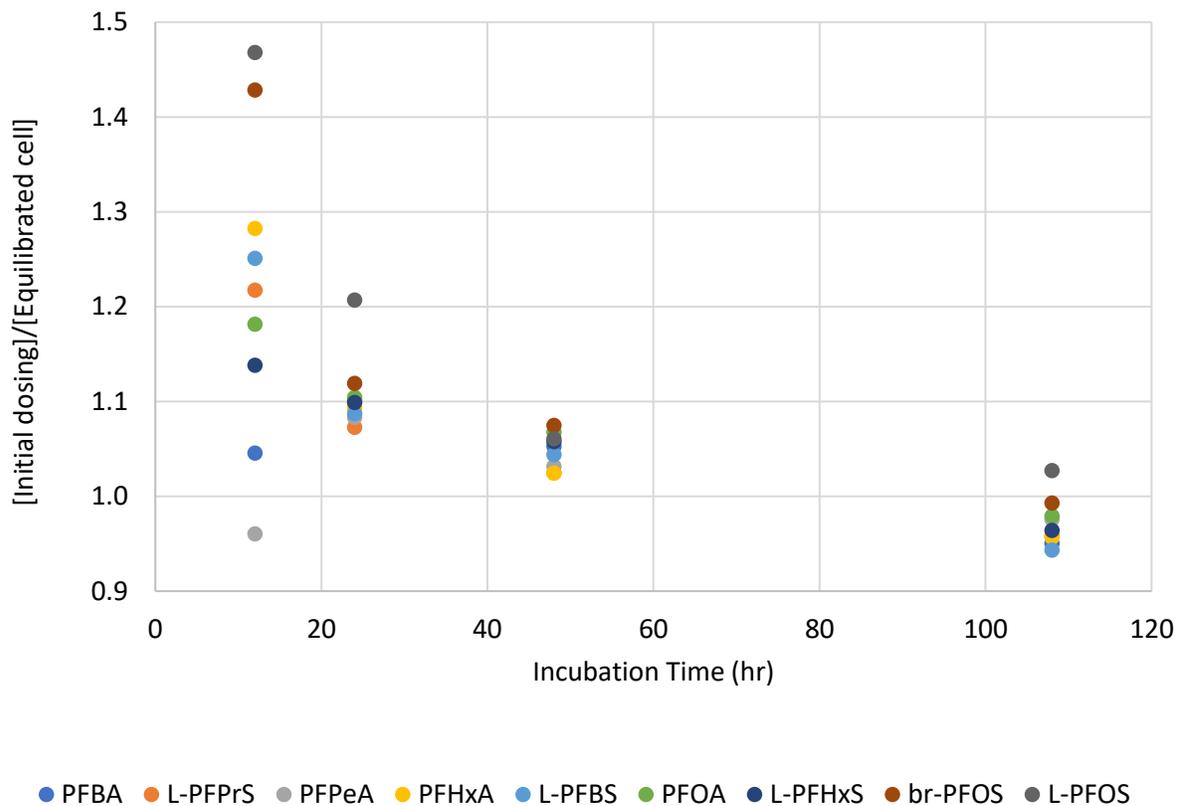
simulated docking result for PFOS and 4E99. The atomic root mean square deviation (RMSD) between the PFOS conformation generated from docking and the experimentally obtained conformation of PFOS in 4E99 was calculated using a pairwise fitting program in PyMol (v2.3.3). A RMSD < 2 Å in this study was considered a successful redocking.

### **B3.2 Details on ligand preparation**

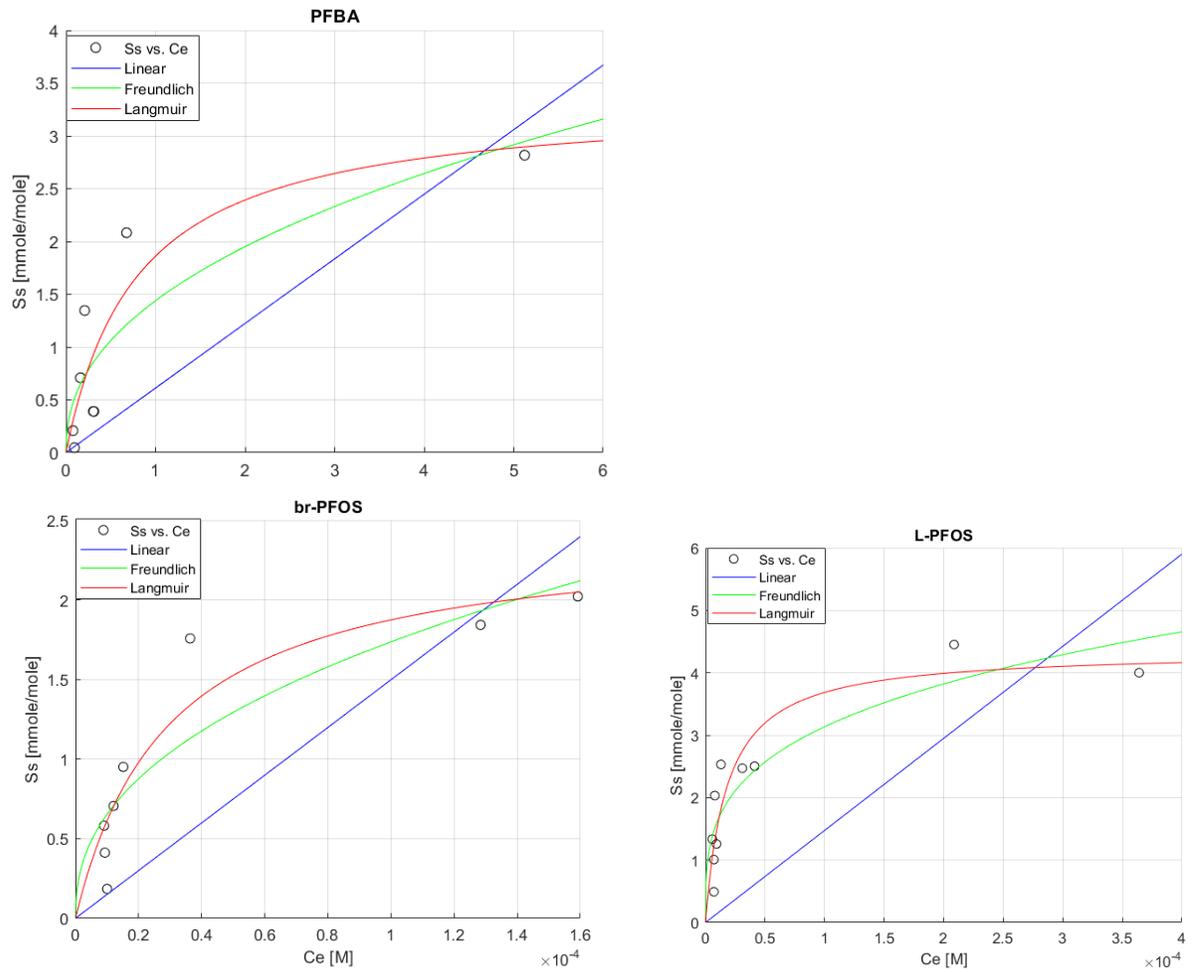
We generated the 3D structures of 26 target PFAS, 18 qualified PFAS, and 18 hydrocarbon surfactants based on their SMILES code (Table S-6.2 and 6.3), using “Generating Conformers” function in DataWarrior V5.2.1 (Sander et al., 2015). We generated the structures (one structure per stereoisomer) with the *Random, low energy* bias algorithm (energy minimized with MMFF94s+ forcefield), which is an optimized energy minimization method that corrected unrealistic torsion parameterization of the original MMFF94s implementation.(Wahl et al., 2019) This process advanced our chance to achieve minimum energy geometry of the ligand, as this optimization process is limited by the accuracy of the starting structures (steepest descent).(Gunasekar et al., 2018) Then we optimized ligand structures through molecular mechanics using MMFF94s forcefield in Avogadro V1.90.0 (Hanwell et al., 2012). Finally, we used UCSF Chimera (v1.13.1) (Pettersen et al., 2004) ‘Prep Dock’ function to generate \*.pdbqt, file type ready for docking.



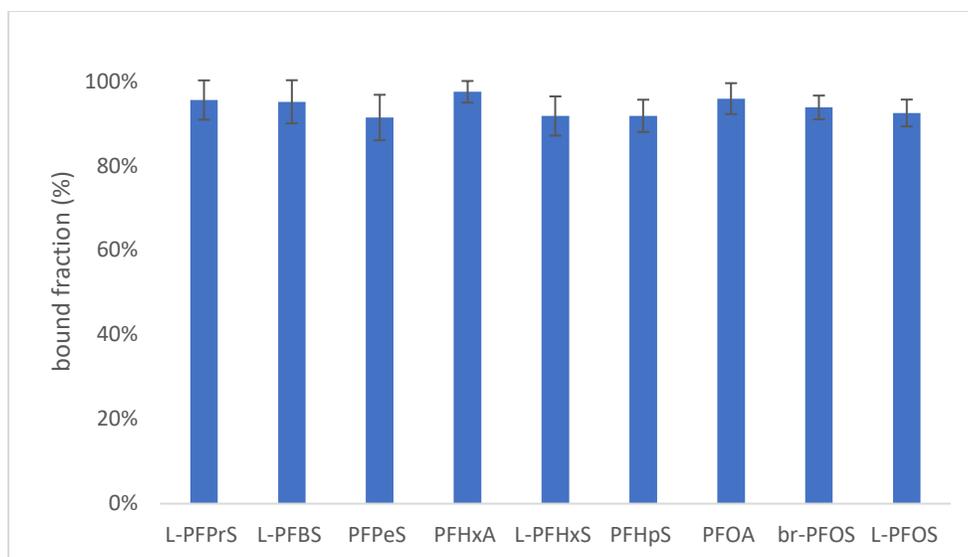
**Figure B1.1** Experimental setup for equilibrium dialysis. For each trial of the experiment ( $n = 4$ ), six measurement cells were prepared to determine association constants. Blank cell was used as a negative control to check contamination and membrane fouling. Control cell was used to check equilibrium and surface adsorption of equilibrium dialysis set-up. The scheme was adapted and adjusted accordingly from the published supplementary information by Allendorf et al. (2019).



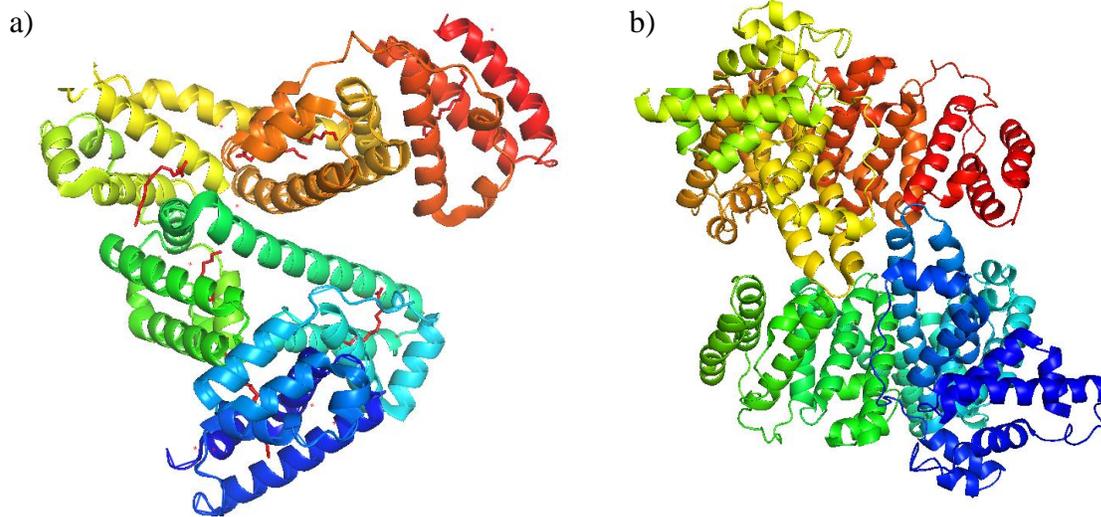
**Figure B1.2** Assessment of equilibration time required for equilibrium dialysis. Equilibrium dialysis was performed with PFAS standards in phosphate buffer saline (PBS) at pH=7.4, 37 °C.



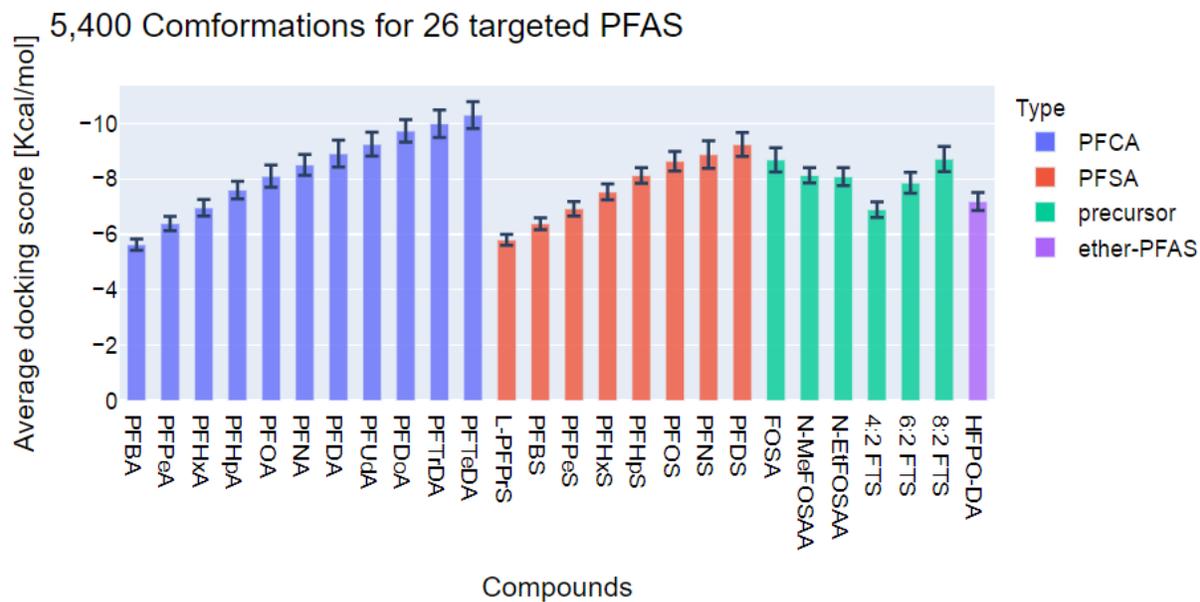
**Figure B1.3** Isotherm models for PFAS adsorbed to HSA. PFBA, L-PFOS and br-PFOS in AFFF dilutions were extracted from the protein aliquot fractions and the aqueous fractions in equilibrium dialysis experiments. Fits of the bound PFAS ( $S_s$ ) versus the free PFAS concentration ( $C_e$ ) are shown for Linear, Freundlich, and Langmuir isotherms.



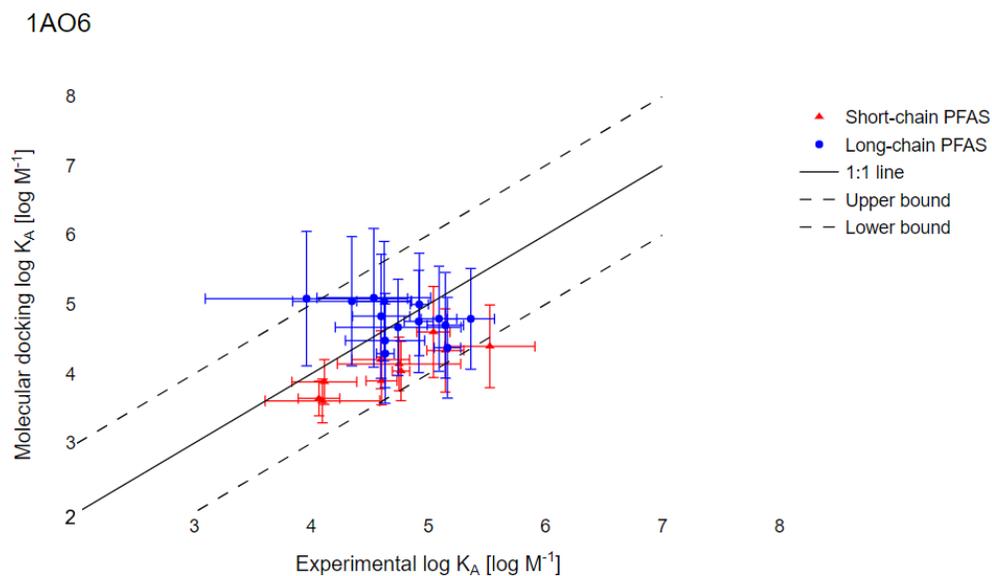
**Figure B1.4** Bound fractions of PFAS compounds in AFFF (df=1E4) to HSA (600  $\mu$ M). The error bar represents one standard deviation for results obtained from three batches.



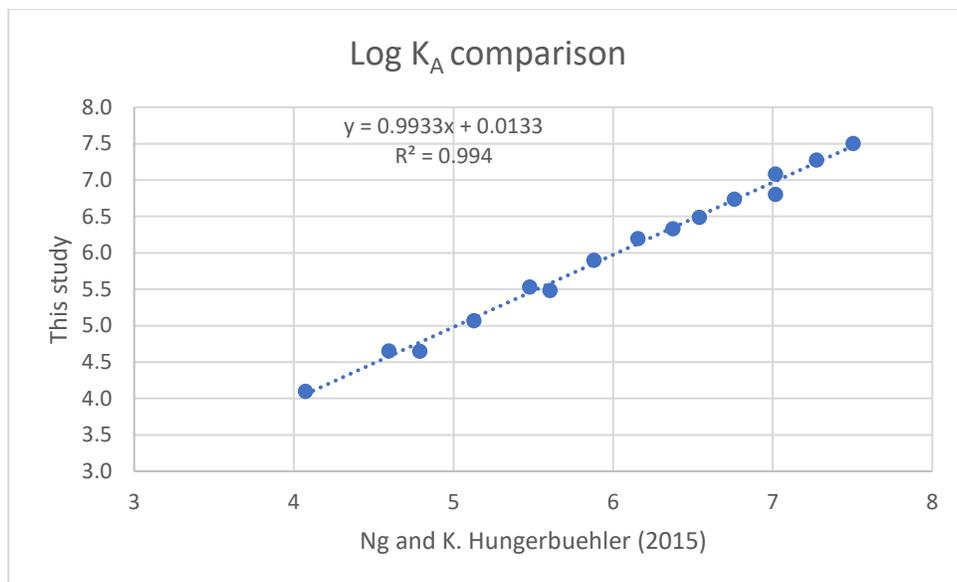
**Figure B2.1** Crystal structures of Human Serum Albumin (HSA). Protein Data Bank (PDB) structures are shown for (a) 1E7G and (b) 1AO6.



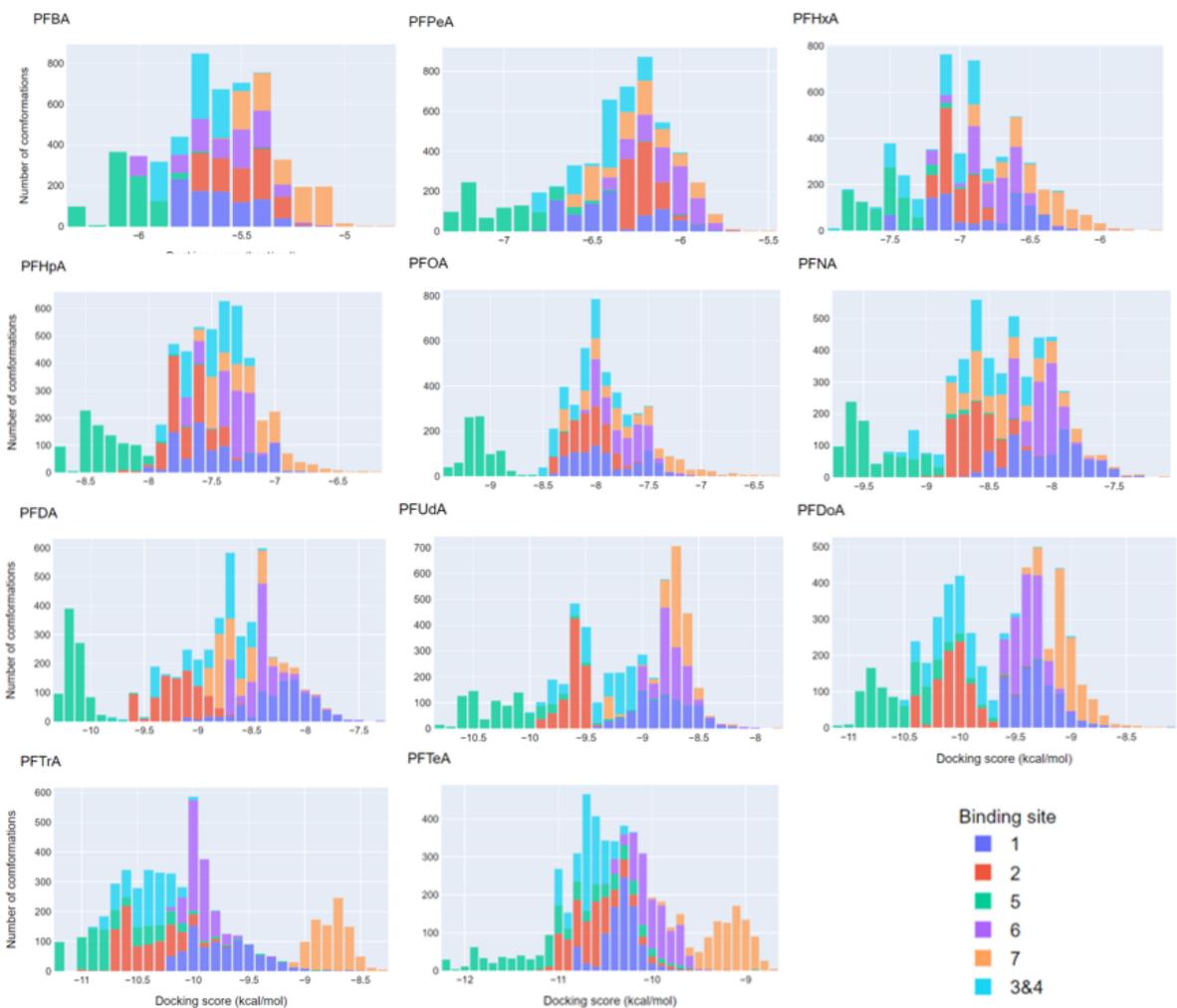
**Figure B2.2** Predicted binding scores of 26-PFAS binding to HSA (1E7G).



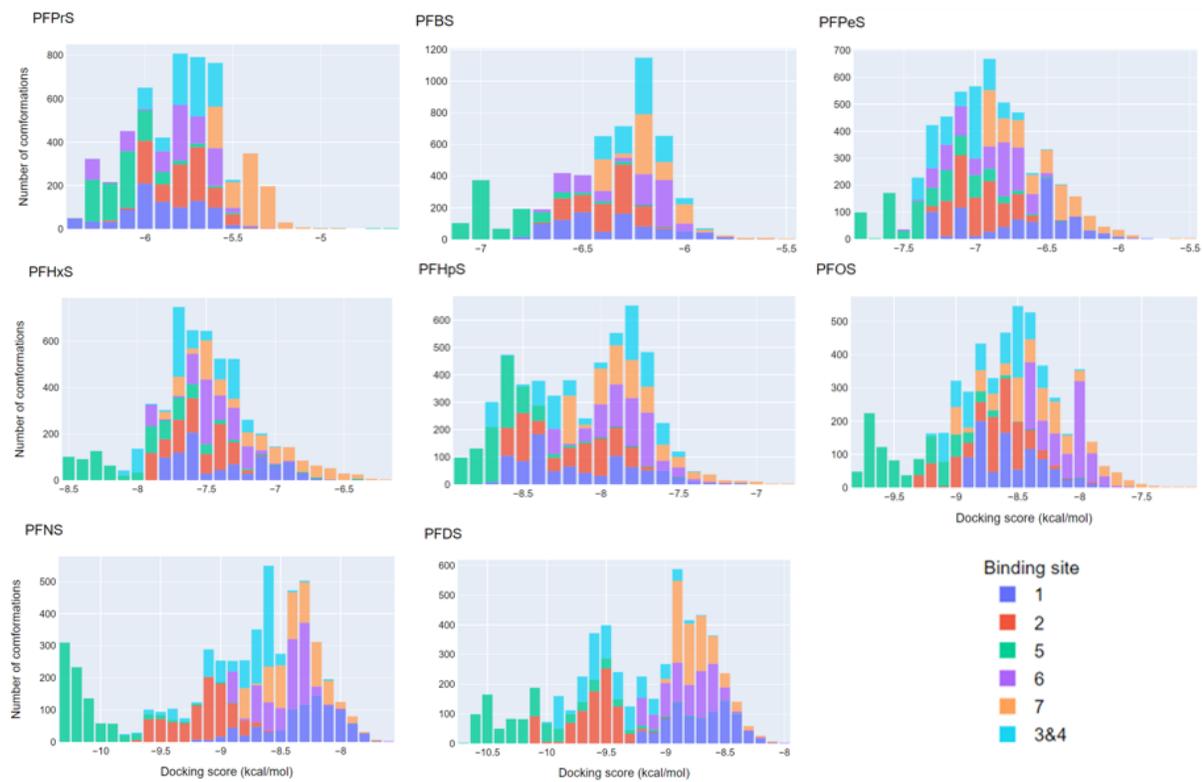
**Figure B2.3.** Molecular docking predicted association constants ( $K_A$ ) of 26 PFAS to HSA. The comparison of experimental  $\log K_A$  with results from molecular simulations was performed with HSA crystal structure 1A06. Solid black lines represent the 1:1 line; dotted lines represent one log unit higher or lower. Error bars reflect one geometric standard deviation (GSD).



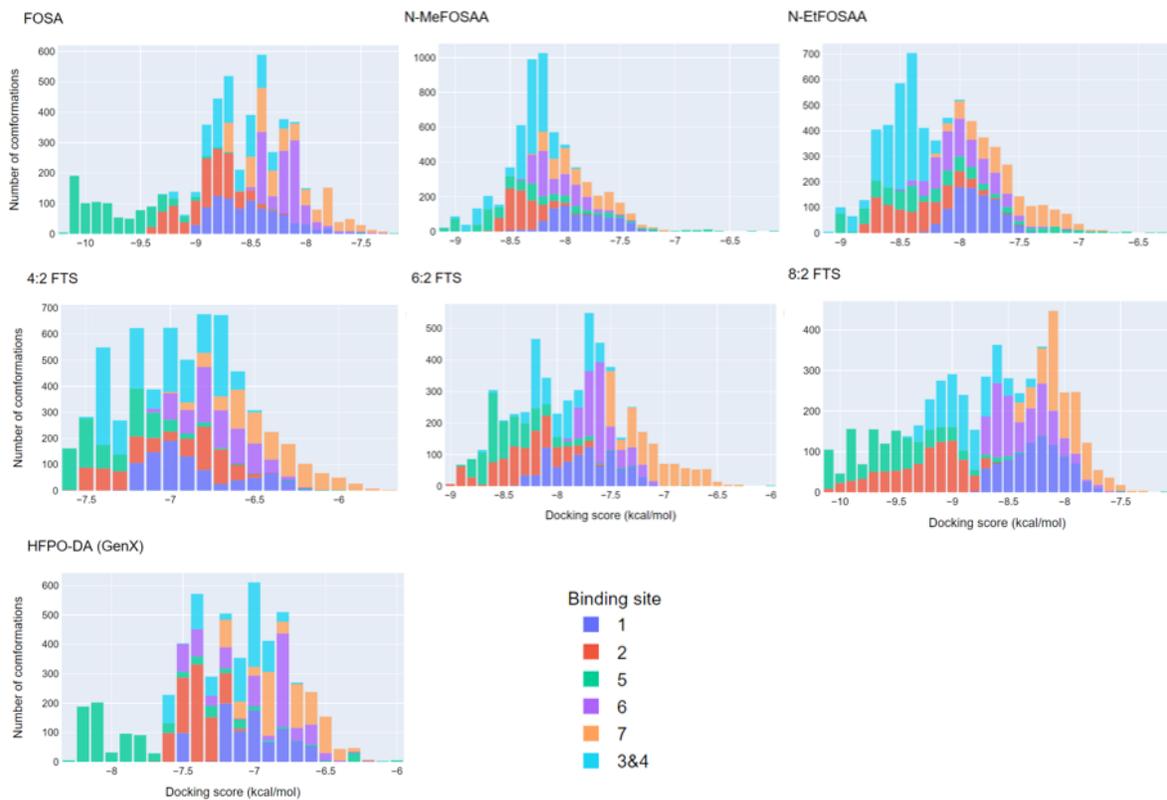
**Figure B2.4** Uncorrected  $K_A$  values for 1E7G from this study correlate to the results reported by Ng and Hungerbuehler (2015). The literature reported values were extracted from the reference Supporting Information Figure S-4(A).



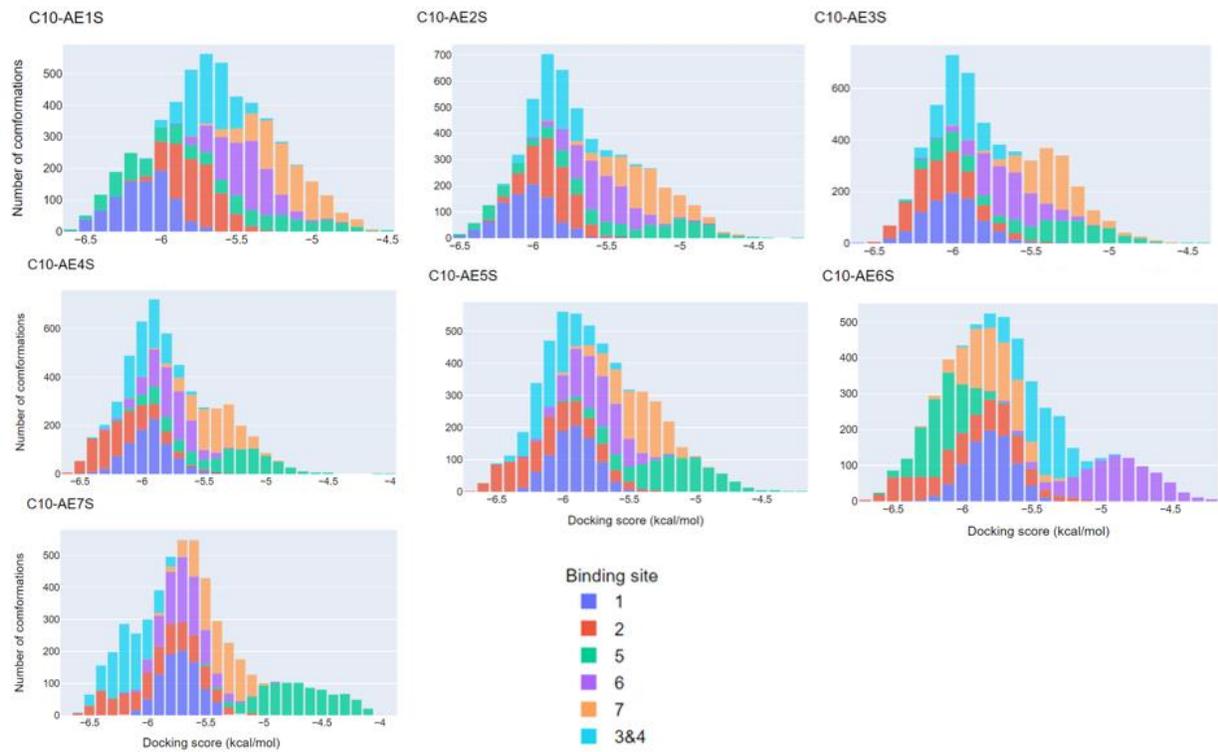
**Figure B3.1** Histograms of molecular docking predicted target PFCA-1E7G binding scores.



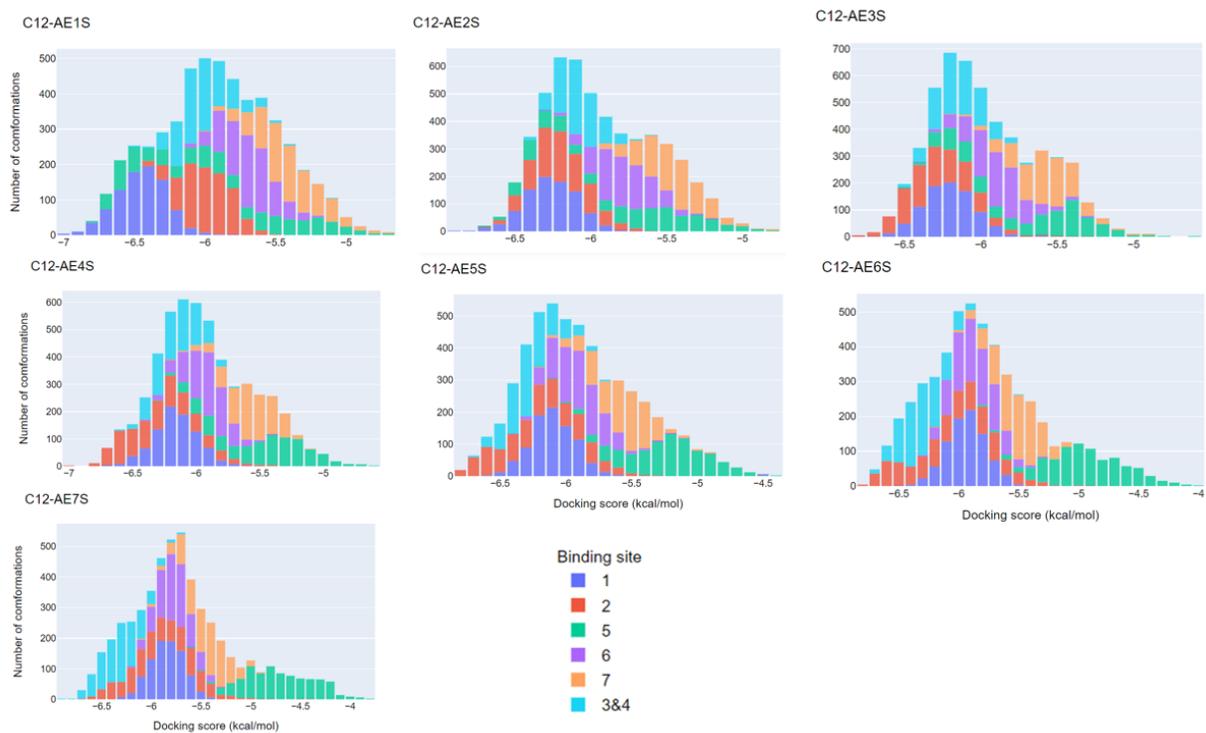
**Figure B3.2** Histograms of molecular docking predicted target PFSA-1E7G binding scores.



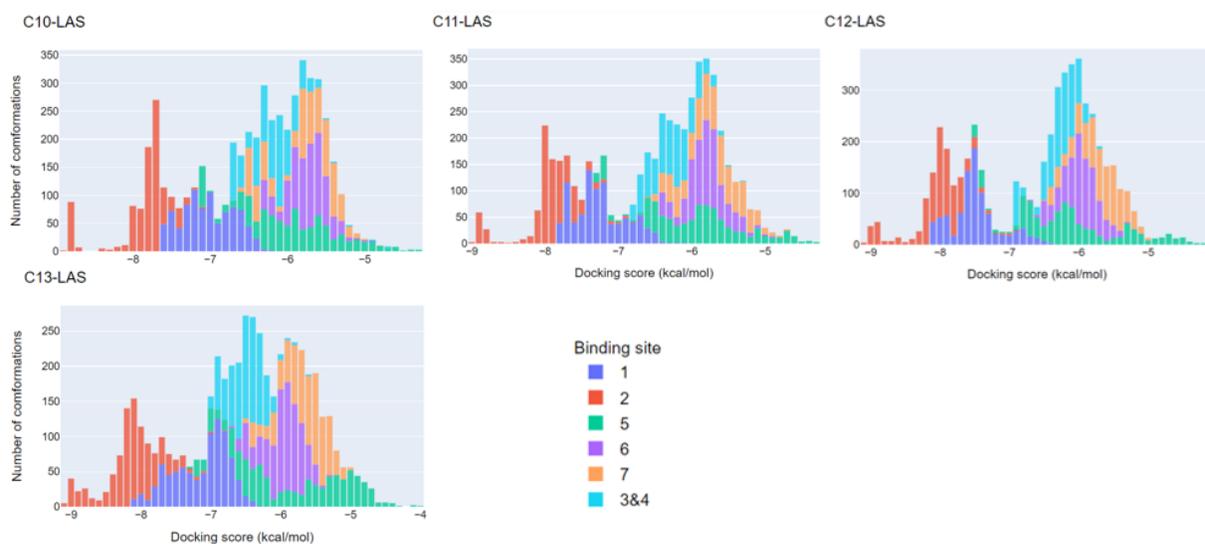
**Figure B3.3** Histograms of molecular docking predicted target PFAS (precursors)-1E7G binding scores.



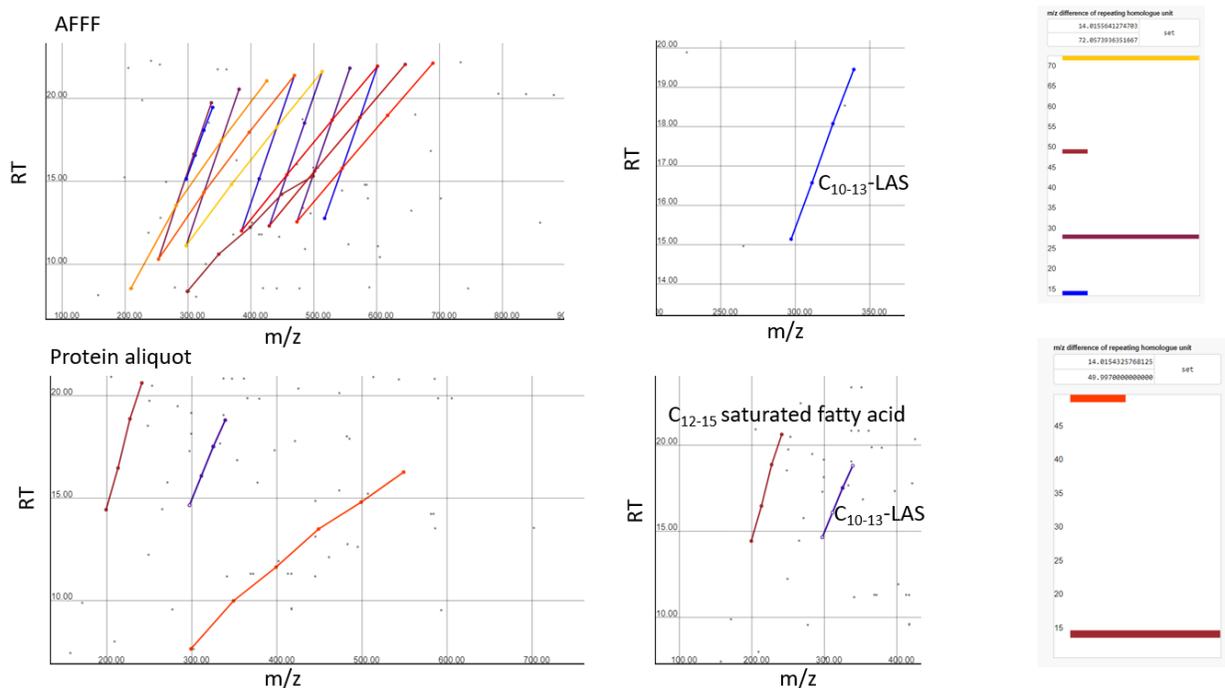
**Figure B3.4** Histograms of molecular docking predicted scores for C10-AEnS binding 1E7G.



**Figure B3.5** Histograms of molecular docking predicted scores for C12-AEnS binding 1E7G.

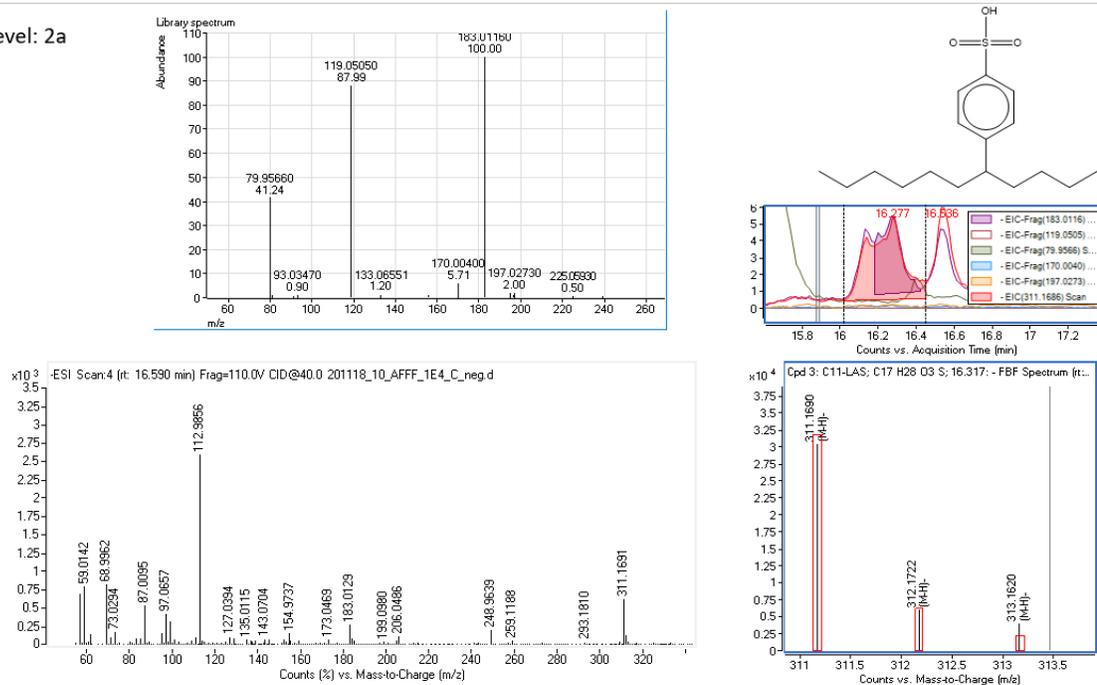


**Figure B3.6** Histograms of molecular docking predicted scores for Cn-LAS binding 1E7G.



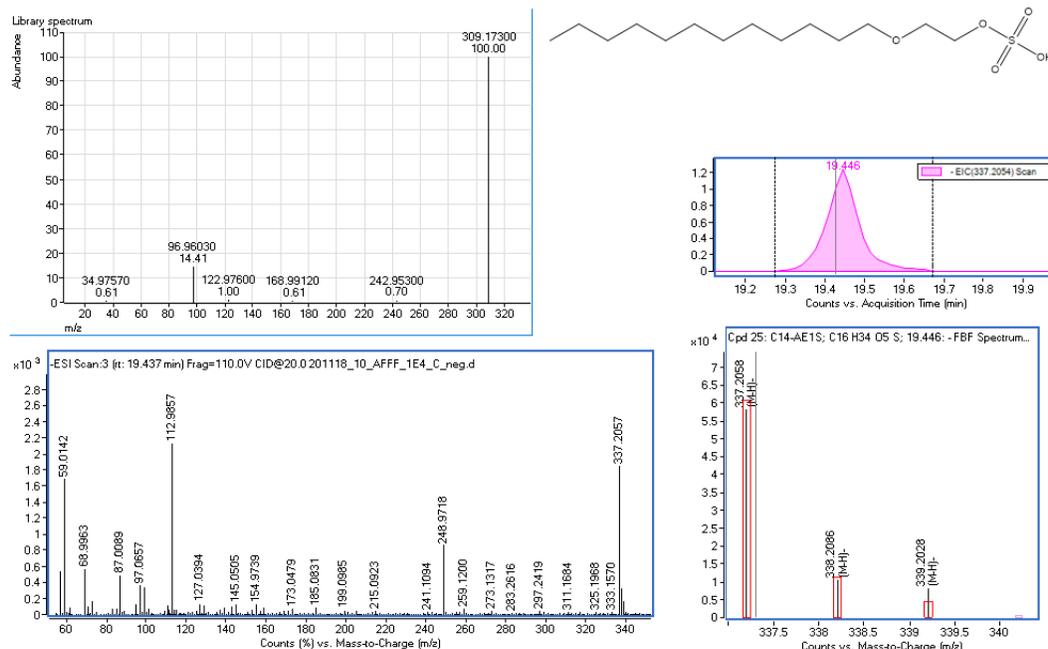
**Figure B4.0** Plot of the retention time (RT) against the mass-to-charge ratio (m/z) in negative ionization mode prepared using the homologous series detection tool EnviHomolog (<http://www.envihomolog.eawag.ch>) across all AFFF dilutions (top) and protein aliquots (bottom). Each point represents a molecular feature; gray points were not identified as part of a homologous series while each colored line connects several points that are part of a homologous series. Colors indicate the mass of the repeating polymeric unit as shown on the inset on the right. Repeating mass-increments of 14.0156 (-CH<sub>2</sub>-), 28.0313 (-C<sub>2</sub>H<sub>4</sub>-), and 44.0262 (-C<sub>2</sub>H<sub>4</sub>O-) were observed in dark red and bright blue colors.

Confidence level: 2a



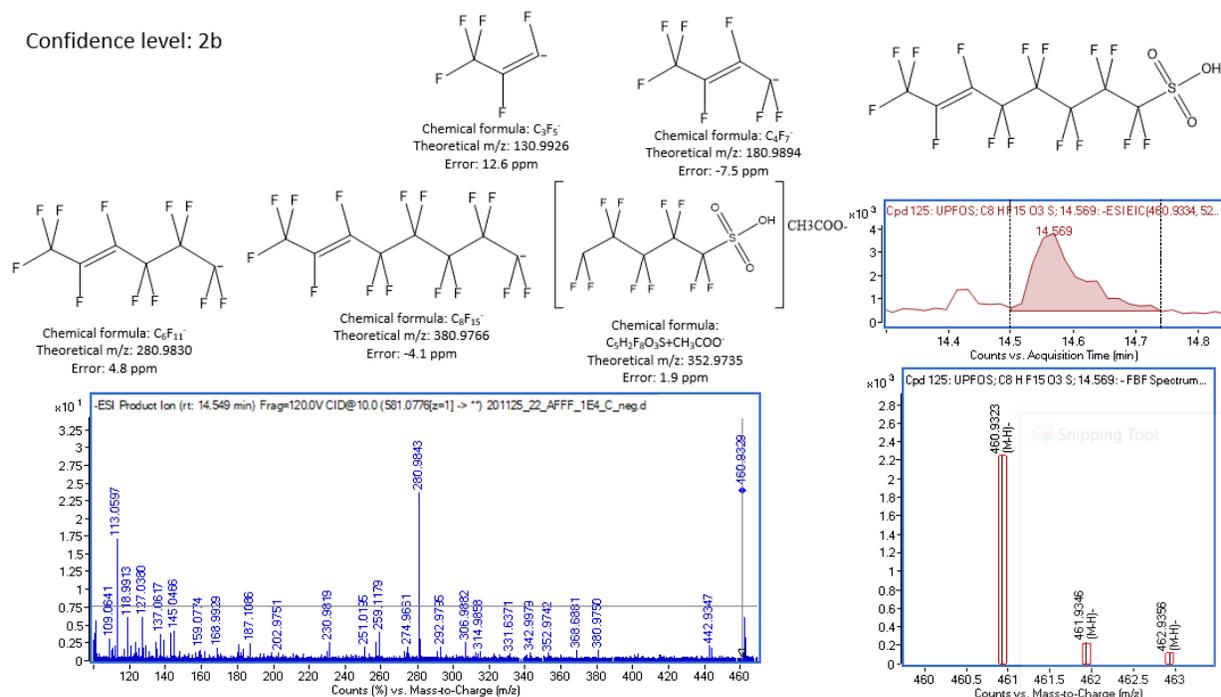
**Figure B4.1** Qualification of Cn-LAS to Confidence Level 2a by matching the MONA library MS/MS spectra in ESI. The lower left figure shows an example of the mass spectra for 40eV collision energy acquisition at the retention time (rt) shown. The top left figure shows the Massbank of North America (MONA) spectra for the suspected compound (shown at top right, C11-LAS acquired at 90eV). The middle right figure shows the chromatogram of fragment ions pulled from the MONA spectra. The lower right corner represents the isotopic pattern for the molecular ion (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red).

Confidence level: 2a



**Figure B4.2** Qualification of Cn-AES to Confidence Level 2a by matching the MONA library MS/MS spectra in ESI. The lower left figure shows an example of the mass spectra for 20eV collision energy acquisition at the retention time (rt) shown. The top left figure shows the Massbank of North America (MONA) spectra for the suspected compound (shown at top right, C12-AES acquired at 60eV). The middle right figure shows the chromatogram of fragment ions pulled from the MONA spectra. The lower right corner represents the isotopic pattern for the molecular ion (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red).

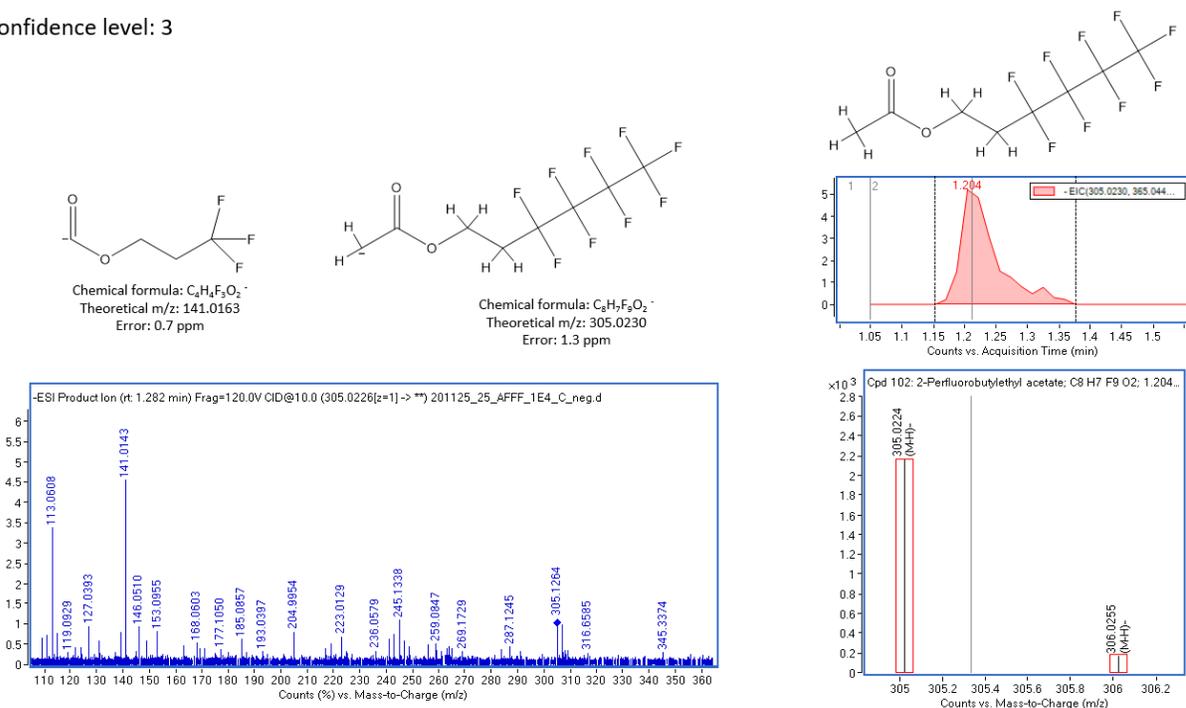
Confidence level: 2b



**Figure B4.3** Qualification of UPFOS to Confidence Level 2b in ESI. Structures for the qualified compound and all fragments ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 10eV collision energy acquisition at the retention time (rt) shown.

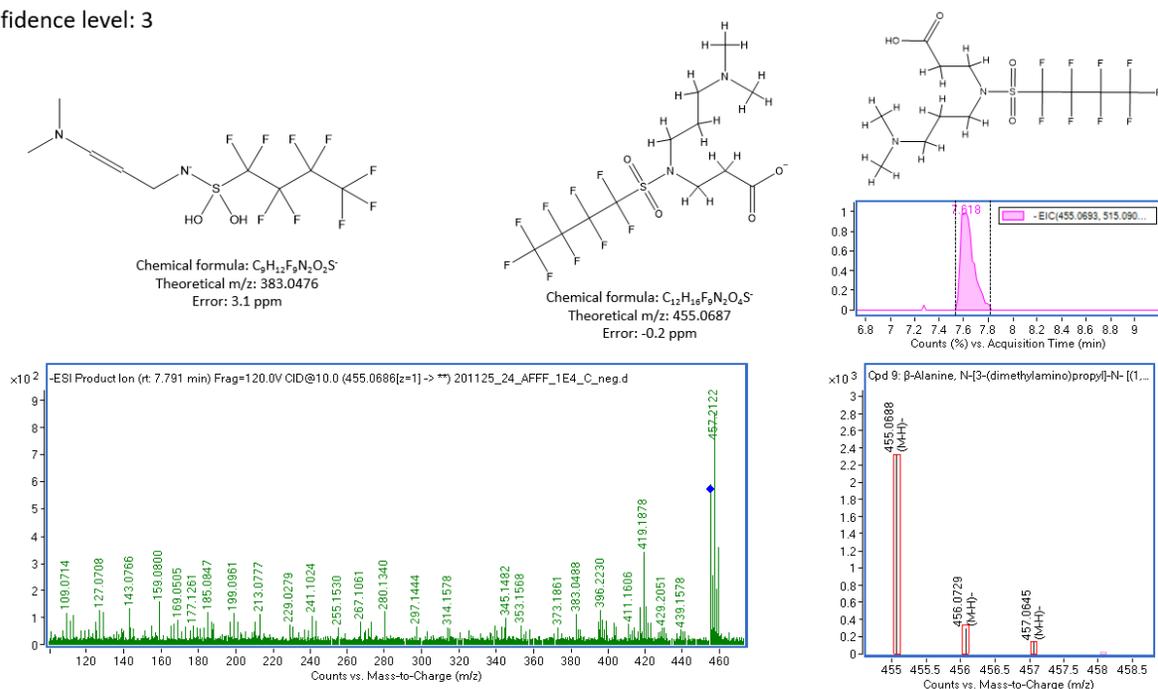


Confidence level: 3



**Figure B4.5** Qualification of N-FBEAc to Confidence Level 3 in ESI. Structures for the qualified compound and all fragments ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 10eV collision energy acquisition at the retention time (rt) shown.

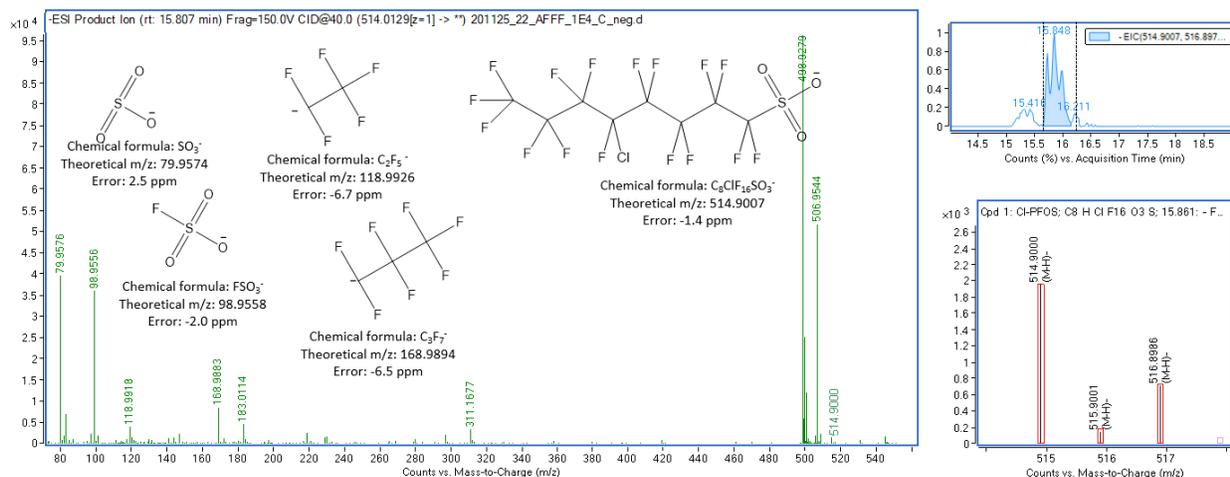
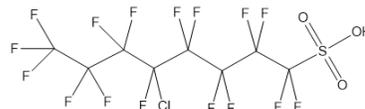
Confidence level: 3



**Figure B4.6** Qualification of N-diMAmP-FBSAP to Confidence Level 3 in ESI. Structures for the qualified compound and all fragments ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 10eV collision energy acquisition at the retention time (rt) shown.



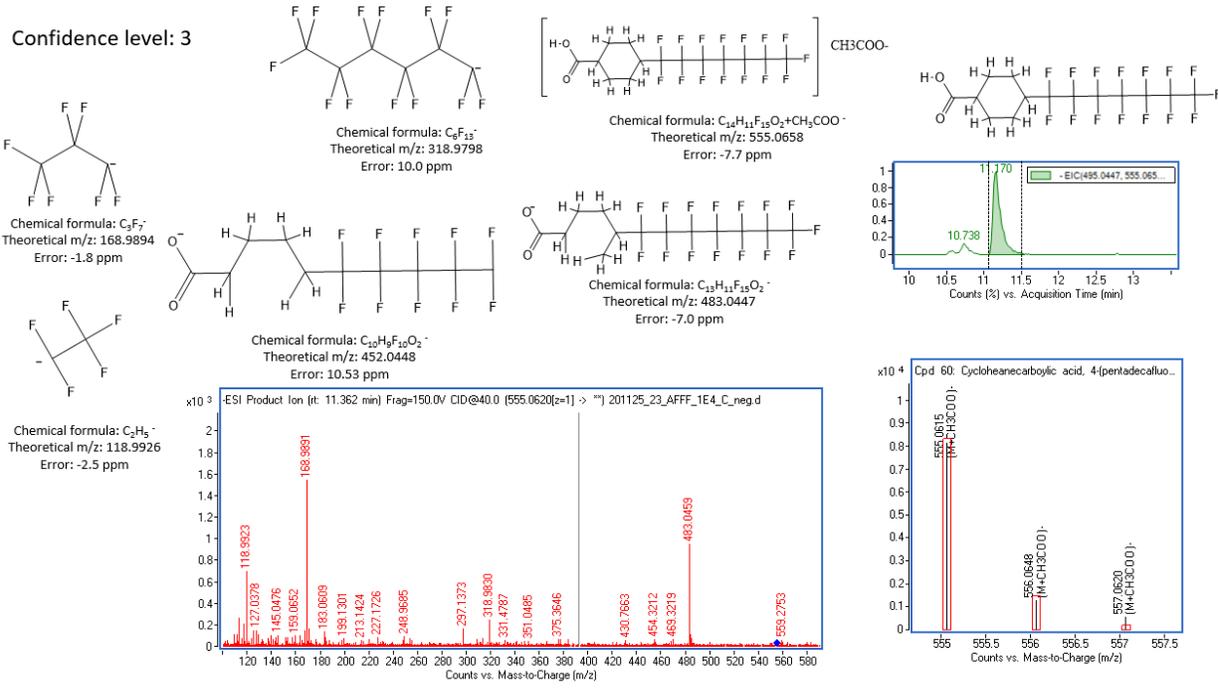
Confidence level: 3



**Figure B4.8** Qualification of Cl-PFOS to Confidence Level 3 in ESI. Structures for the qualified compound and all fragments ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 40eV collision energy acquisition at the retention time (rt) shown.

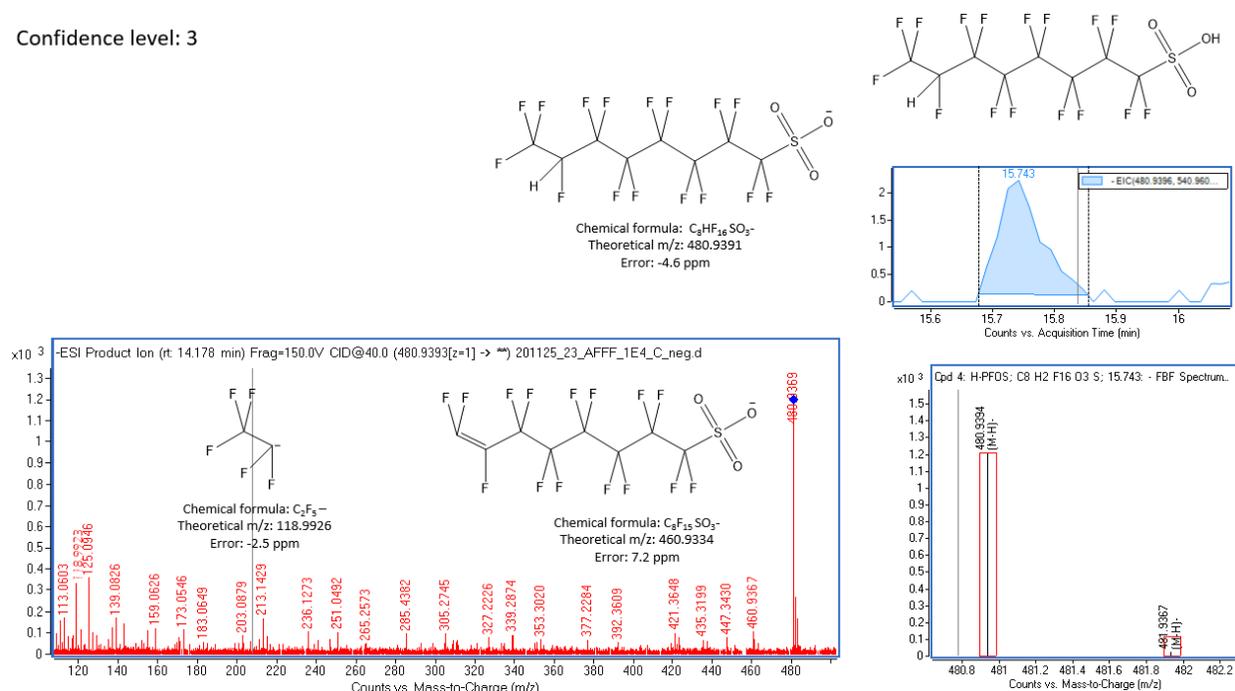




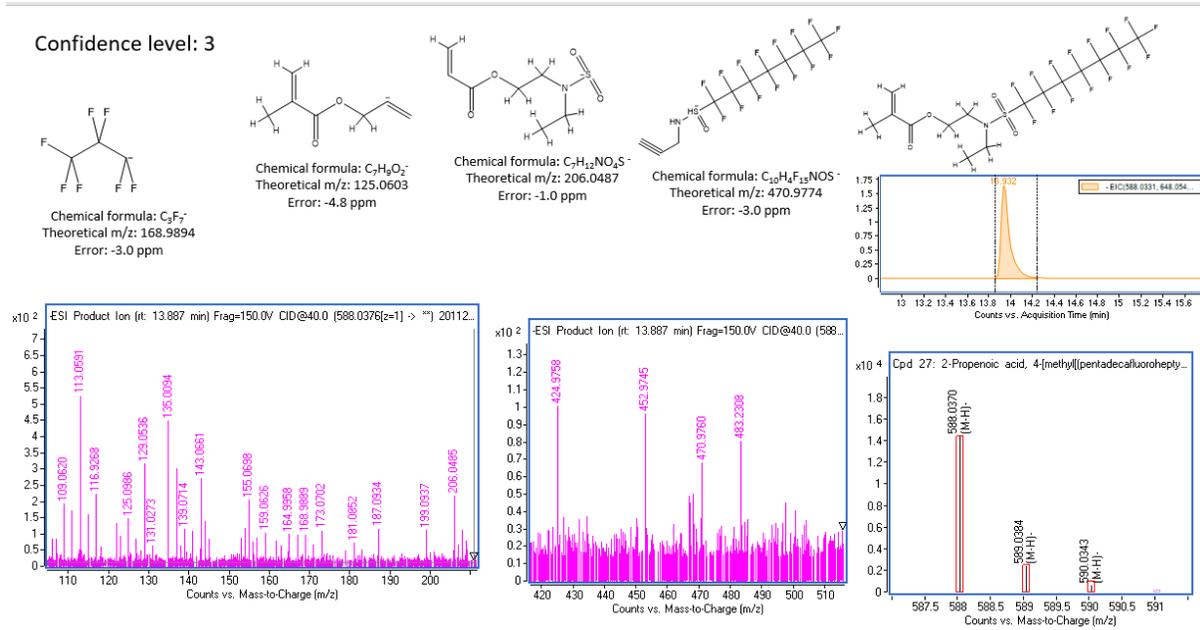


**Figure B4.11** Qualification of 4-FHp-CycHxA to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 40eV collision energy acquisition at the retention time (rt) shown.

Confidence level: 3

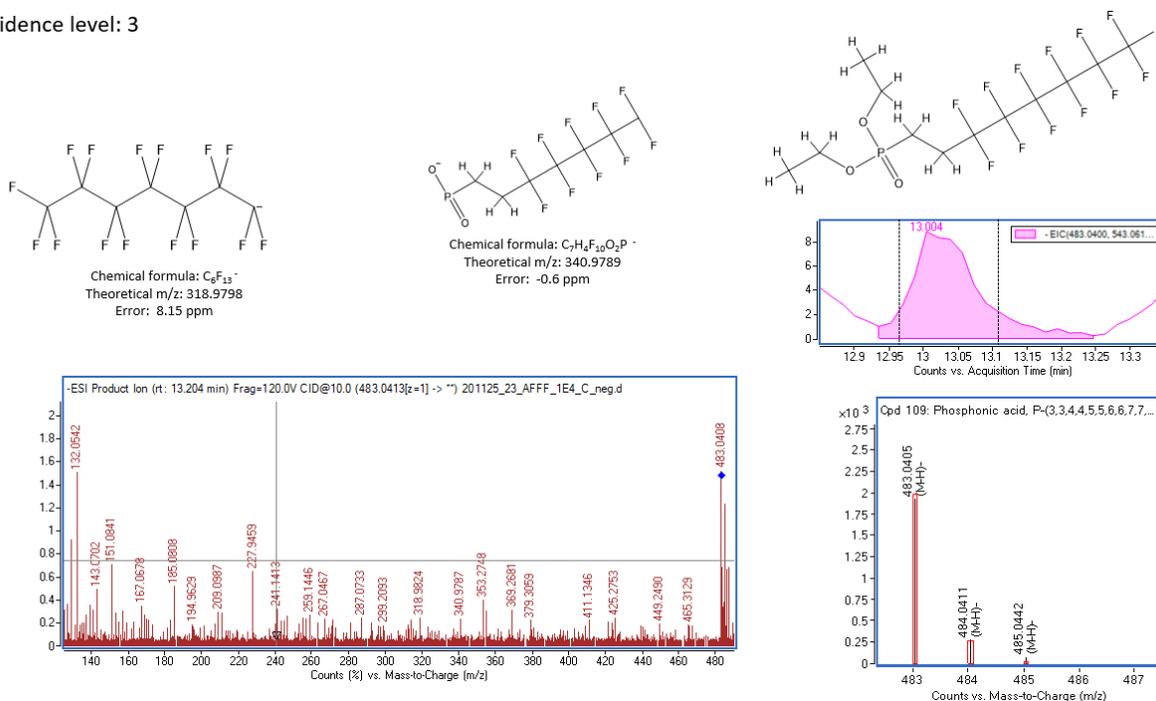


**Figure B4.12** Qualification of H-PFDS to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 40eV collision energy acquisition at the retention time (rt) shown.



**Figure B4.13** Qualification of N-EFHpSA-EMAC to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 40eV collision energy acquisition at the retention time (rt) shown.

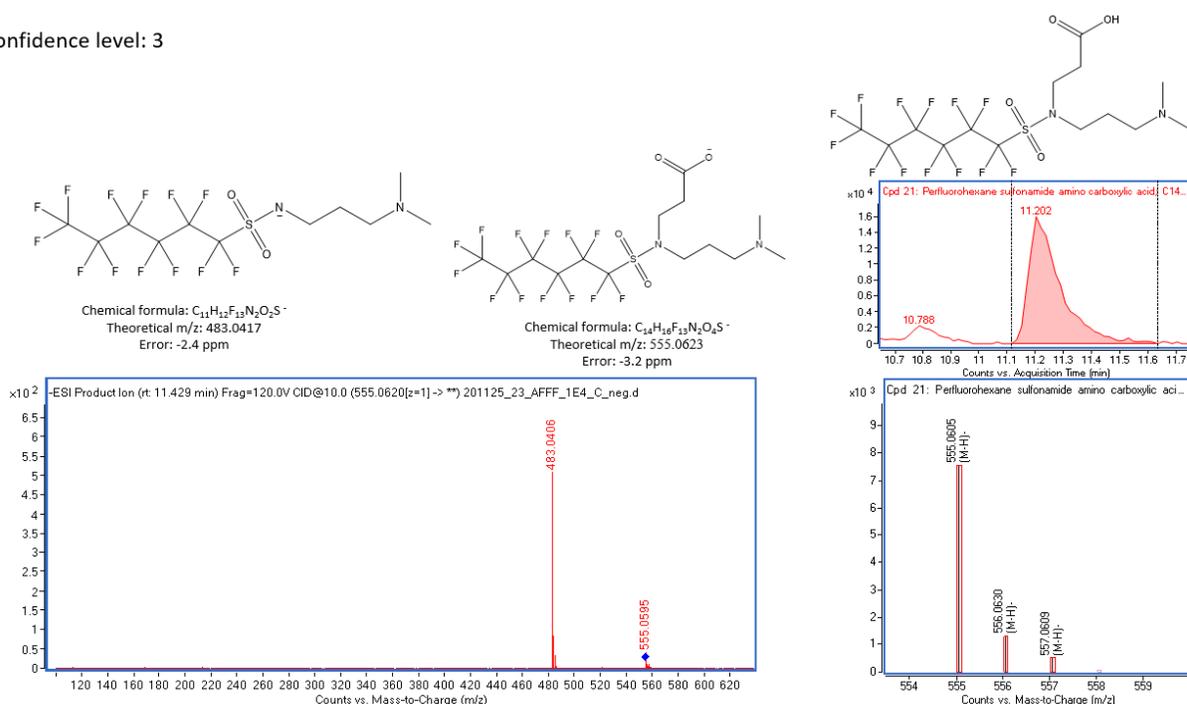
Confidence level: 3



**Figure B4.14** Qualification of 8:2 monoPAP-diEes to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 10eV collision energy acquisition at the retention time (rt) shown.

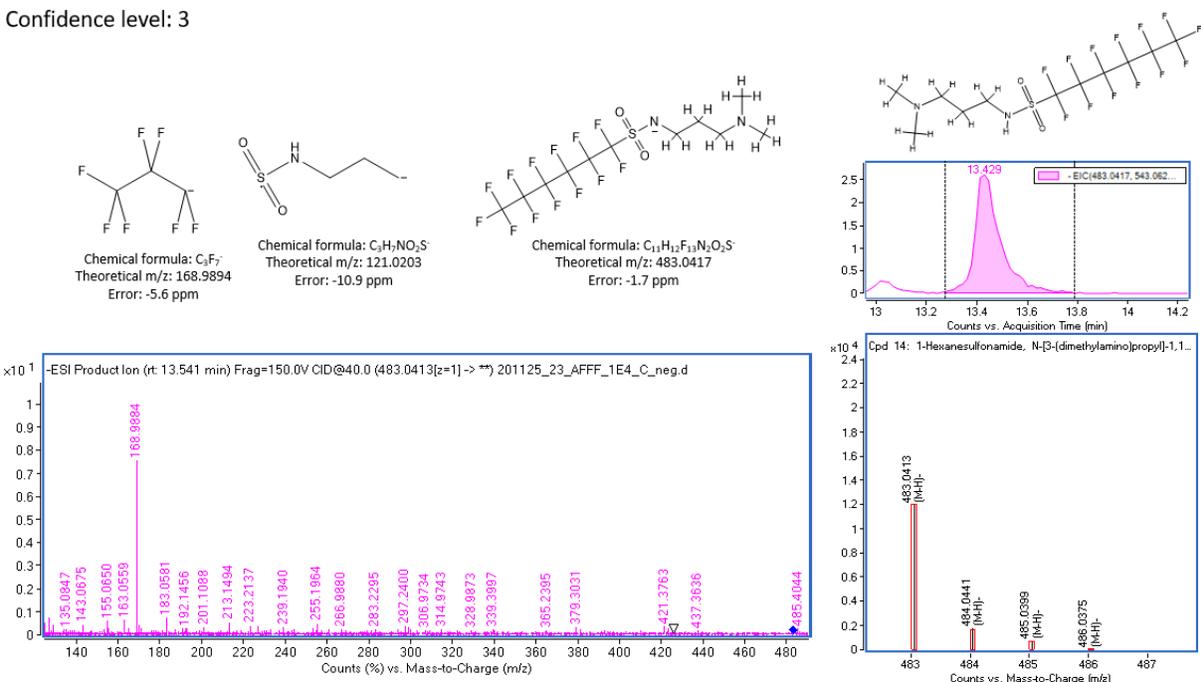


Confidence level: 3



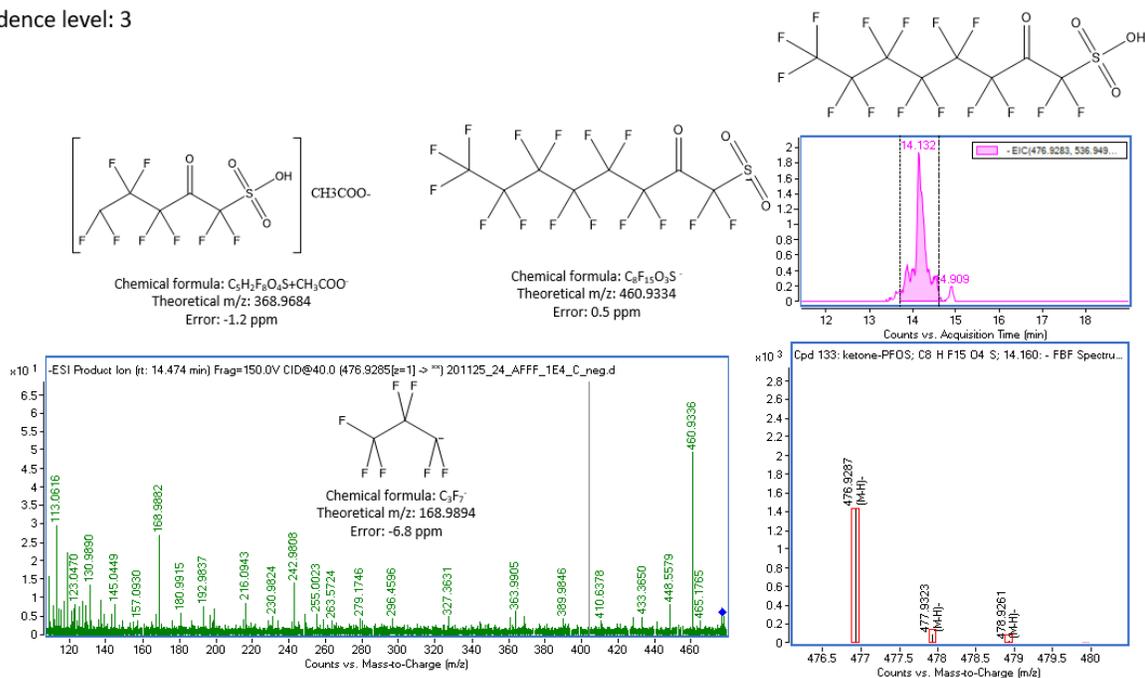
**Figure B4.16** Qualification of N-TMAmP-FHxSAE to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 10eV collision energy acquisition at the retention time (rt) shown.

Confidence level: 3



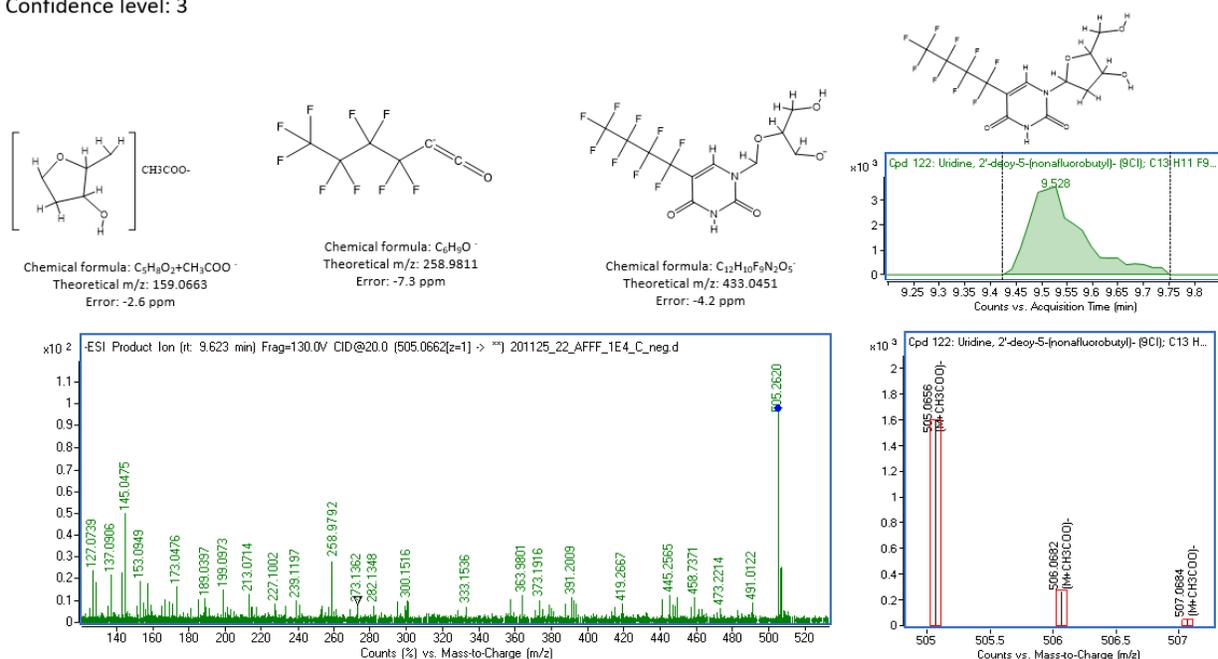
**Figure B4.17** Qualification of N-diMAMp-FHxSA to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 40eV collision energy acquisition at the retention time (rt) shown.

Confidence level: 3



**Figure B4.18** Qualification of K-PFOS to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample (df = 1E4) at 40eV collision energy acquisition at the retention time (rt) shown.

Confidence level: 3



**Figure B4.19** Qualification of Uridine-FB to Confidence Level 3 in ESI. Structures for the qualified compound and all fragment ions detected are shown. The middle right inset shows the extracted chromatogram for the molecular ion. The lower right inset figure shows the isotopic pattern (in black) in comparison to the theoretical isotopic pattern for the molecular formula (in red). The lower left figure shows an example of the mass spectra for the targeted MS/MS of the AFFF sample ( $df = 1E4$ ) at 20eV collision energy acquisition at the retention time (rt) shown.

**Table B1.1** Mass (ng) of PFAS present in each fraction of equilibrium dialysis extracts.

PFAS dosage	Covalent			Non-covalent			Aqueous portion		
	PFBS	PFOA	PFOS	PFBS	PFOA	PFOS	PFBS	PFOA	PFOS
26 PFAS [200 ng]	3.91	20.25	3.70	110.95	51.50	206.63	11.0	3.1	6.8
26 PFAS [80 ng]	6.12	16.10	3.64	52.19	50.39	61.77	6.4	1.8	6.3
26 PFAS [40 ng]	8.29	3.57	3.26	21.00	28.26	31.25	6.6	2.0	7.7
26 PFAS [20 ng]	13.22	7.43	3.20	13.40	18.74	8.77	<LOQ	1.4	15.3
26 PFAS [10 ng]	8.22	4.36	3.27	4.81	9.73	6.11	7.2	2.6	7.3
26 PFAS [4 ng]	2.61	0.95	3.16	4.89	11.19	2.57	<LOQ	<LOQ	6.5
AFFF (df=2e3)	2.43	17.34	4.50	9.69	10.20	584.25	6.2	2.3	6.4
AFFF (df=4e3)	1.18	12.95	3.87	9.98	8.96	267.56	<LOQ	<LOQ	6.5
AFFF (df=1.6e4)	1.72	15.10	3.25	5.36	8.68	121.98	<LOQ	<LOQ	7.8
AFFF (df=2e4)	4.56	20.42	3.54	9.50	4.24	95.45	<LOQ	0.5	7.2
AFFF (df=4e4)	0.88	22.25	3.27	5.85	7.88	53.11	<LOQ	<LOQ	7.1
AFFF (df=8e4)	0.90	21.84	3.35	5.41	9.87	16.10	<LOQ	<LOQ	7.1

**Table B2.1** Molecular docking searching information for 1E7G

Binding Sites	Search Center(Å)			Size of Search Box(Å)		
	x	y	z	x	y	z
FA1	32.913	15.789	32.677	18	20	30
FA2	47.386	9.578	20.034	20	24	22
FA3/4	11.193	5.903	20.079	22	28	22
FA5	0.702	5.791	42.116	20	20	22
FA6	24.958	9.578	1.116	20	22	22
FA7	33.244	10.42	15.488	20	22	20

**Table B2.2** Molecular docking searching information for 1AO6

Binding Sites	Search Center(Å)			Size of Search Box(Å)		
	x	y	z	x	y	z
FA1	45	25	18	28	28	34
FA2	50	45	18	32	25	33
FA3/4	20	30	20	20	35	25
FA5	18	20	0	28	20	25
FA6	10	40	45	23	25	25
FA7	35	35	40	30	25	25

**Table B2.3** Molecular docking searching information for 4E99

Binding Sites	Search Center(Å)			Size of Search Box(Å)			RMSD range for 9 poses (Å)
	x	y	z	x	y	z	
FA3/4	-1	5	34	15	15	12	0.207-1.973
FA6	-18	-12	18	20	15	18	0.536-1.915

**Table B5.1** LC-QTOF-MS acquisition method (LC-QTOF-MS acquisition)

LC-QTOF-MS Method		
Injection Volume		10 $\mu$ L
LC Settings		
Mobile Phases	A (neg)	H <sub>2</sub> O + 20 mM Ammonium Acetate
	B (neg)	acetonitrile
Solvent Flow		0.40 mL/min
Gradient for samples		25% B for 1.5 min
		25%-90% B in 25 min
		100% B for 5 min
		equilibration to initial conditions for 3 min
Gradient for blanks		2% B for 1.5 min
		0%-100% B in 23 min
		100% B for 5 min
		equilibration to initial conditions for 3 min
Column		ZORBAX RRHD Eclipse Plus 95 $\text{\AA}$ , 2.1 x 150 mm, 1.8 $\mu$ m)
Column Temperature		30 $^{\circ}$ C
Gas Temperature		300 $^{\circ}$ C
Drying Gas Flow		12 l/min
Nebulizer		25 psig
Sheath Gas Temperature		350 $^{\circ}$ C
Sheath Gas Flow		11 l/min
Vcap		3000 (neg)
Fragmentor		110 V
scan range		50-1200 m/z
scan speed		4 spectra/s
All-Ions Acquisition		Collision Energy (CE): 0, 10, 20, 40
Reference Mass Correction		neg: masses 112.9855, 1033.9881

**Table B5.2** Suspect screening software search algorithm criteria for LC-QTOF-MS acquired

data

Suspect Screening LC-QTOF-MS	
Software	Agilent MassHunter Qualitative Analysis (B.08.00)
Workflow	Find Compounds by Formula
Values to Match	Mass
Libraries	In-house PFASs database (3794 PFASs, 91 compounds with MS/MS spectra from in-house standard, massbank, and CFM-ID)
Extraction Algorithm	Agile 2
Match Tolerance Masses	$\pm 10$ ppm
Allowed Adducts	[M-H] <sup>-</sup> [M-CH <sub>3</sub> OOH] <sup>-</sup>
Allowed Charge State	1
Isotope Model	common organic molecules
Peak spacing tolerance	0.0025 m/z, plus 7 ppm
Scoring (Weight)	Mass Score: 100
	Isotope Abundance Score: 50
	Isotope Spacing Score: 60
Find by Formula Score	>80 (out of 100)
Absolute Height	> 1000 counts
Confirm with fragment ions	if MS/MS spectra available
number of most specific ions from MS/MS library	5
RT difference	$\pm 0.2$ min
S/N ratio	not applied
Coelution score	> 85%
Minimum Number of Qualified Fragments	1

**Table B5.3** <sup>19</sup>F and <sup>1</sup>H NMR acquisition method

Parameter	Value
Solvent	MeOD for qNMR; D2O for protein titrations
Temperature	299.9
Pulse Sequence	zgflqn
Experiment	1D
Probe	5 mm PABBO BB/ 19F-1H/ D Z-GRD Z109128/ 0012
Number of Scans	2048 for solutions with one PFAS and 4096 for mixtures of PFASs
Receiver Gain	1149.4
Relaxation Delay	2
Pulse Width	19
Presaturation Frequency	
Acquisition Time	0.6947
Class	
Spectrometer Frequency	470.53
Spectral Width	94339.6
Lowest Frequency	-88885.6
Nucleus	<sup>19</sup> F
Acquired Size	65536
Spectral Size	131072
•Acquire <sup>1</sup> H	•Acquire <sup>19</sup> F NMR
•Spectra window: 12 ppm to -2 ppm	•Spectra window: 0 ppm to -200 ppm
•Relaxation time: 2s	•Relaxation time: 2s
•Number of scan: 10	•Number of scan: 2048
•Dwell time 5.3 usec	•Dwell time 17.7 usec

**Table B6.1** List of LC-QTOF-MS targeted compounds, internal standards, LOQ, and extraction recoveries

Chemical Name	Acronym	Neutral Molecular Formula	Internal Standard	Limit of quantifications [ng/mL]	Absolute recoveries for HSA extractions (%) <sup>*</sup>
Perfluoro-n-butanoic acid	PFBA	C <sub>4</sub> H <sub>0</sub> F <sub>7</sub>	<sup>13</sup> C <sub>4</sub> -PFBA	0.2	69±36%
Perfluoro-n-pentanoic acid	PFPeA	C <sub>5</sub> H <sub>0</sub> F <sub>9</sub>	<sup>13</sup> C <sub>5</sub> -PFPeA	0.05	129±29%
Perfluoro-n-hexanoic acid	PFHxA	C <sub>6</sub> H <sub>0</sub> F <sub>11</sub>	<sup>13</sup> C <sub>5</sub> -PFHxA	0.05	
Perfluoro-n-heptanoic acid	PFHpA	C <sub>7</sub> H <sub>0</sub> F <sub>13</sub>	<sup>13</sup> C <sub>4</sub> -PFHpA	0.05	93±5%
Perfluoro-n-octanoic acid	PFOA	C <sub>8</sub> H <sub>0</sub> F <sub>15</sub>	<sup>13</sup> C <sub>8</sub> -PFOA	0.05	86±22%
Perfluoro-n-nonanoic acid	PFNA	C <sub>9</sub> H <sub>0</sub> F <sub>17</sub>	<sup>13</sup> C <sub>9</sub> -PFNA	0.05	108±13%
Perfluoro-n-decanoic acid	PFDA	C <sub>10</sub> H <sub>0</sub> F <sub>19</sub>	<sup>13</sup> C <sub>6</sub> -PFDA	0.5	73±22%
Perfluoro-n-undecanoic acid	PFUdA	C <sub>11</sub> H <sub>0</sub> F <sub>21</sub>	<sup>13</sup> C <sub>7</sub> -PFUdA	0.05	50±47%
Perfluoro-n-dodecanoic acid	PFDoA	C <sub>12</sub> H <sub>0</sub> F <sub>23</sub>	<sup>13</sup> C- PFDoA	0.05	82±14%
Perfluoro-n-tridecanoic acid	PFTrDA	C <sub>13</sub> H <sub>0</sub> F <sub>25</sub>	<sup>13</sup> C <sub>2</sub> - PFTeDA	0.5	101±1%
Perfluoro-n-tetradecanoic acid	PFTeDA	C <sub>14</sub> H <sub>0</sub> F <sub>27</sub>	<sup>13</sup> C <sub>2</sub> - PFTeDA	0.5	123±7%
Perfluoro-1-propanesulfonic acid	PFPPrS	C <sub>3</sub> H <sub>0</sub> SF <sub>7</sub> <sup>*</sup>	<sup>13</sup> C <sub>3</sub> - PFBS	4	82±12%
Perfluoro-1-butanesulfonic acid	PFBS	C <sub>4</sub> H <sub>0</sub> SF <sub>9</sub>	<sup>13</sup> C <sub>3</sub> - PFBS	0.05	84±15%
Perfluoro-1-pentanesulfonic acid	PFPeS	C <sub>5</sub> H <sub>0</sub> SF <sub>11</sub>	<sup>13</sup> C <sub>5</sub> - PFHxA	0.05	77±30%
Perfluorohexanesulfonic acid	PFHxS	C <sub>6</sub> H <sub>0</sub> SF <sub>13</sub>	<sup>13</sup> C <sub>3</sub> - PFHxS	0.5	119±12%
Perfluoroheptanesulfonic acid	PFHpS	C <sub>7</sub> H <sub>0</sub> SF <sub>15</sub>	<sup>13</sup> C <sub>3</sub> - PFHxS	0.05	108±14%
Linear perfluorooctanesulfonic acid	L-PFOS	C <sub>8</sub> H <sub>0</sub> SF <sub>17</sub> <sup>**</sup>	<sup>13</sup> C <sub>8</sub> - PFOS	0.025	92±11%
Branched perfluorooctanesulfonic acid	br-PFOS	C <sub>8</sub> H <sub>0</sub> SF <sub>17</sub> <sup>***</sup>	<sup>13</sup> C <sub>8</sub> - PFOS	2.5	129±16%
Perfluorononanesulfonic acid	PFNS	C <sub>9</sub> H <sub>0</sub> SF <sub>19</sub>	<sup>13</sup> C <sub>8</sub> - PFOS	0.05	105±16%
Perfluorodecanesulfonic acid	PFDS	C <sub>10</sub> H <sub>0</sub> SF <sub>21</sub>	<sup>13</sup> C <sub>8</sub> - PFOS	0.05	80±15%

Perfluoro-1-octanesulfonamide	FOSA	C <sub>8</sub> H <sub>2</sub> O <sub>2</sub> NSF <sub>17</sub>	13C <sub>8</sub> -PFOS	0.5	134±14%
N-methylperfluoro-1-octanesulfonamide acetic acid	MeFOSAA	C <sub>11</sub> H <sub>6</sub> F <sub>17</sub> NO <sub>4</sub> S	d <sub>3</sub> -MeFOSAA	0.5	129±5%
N-ethylperfluoro-1-octanesulfonamide acetic acid	EtFOSAA	C <sub>12</sub> H <sub>8</sub> F <sub>17</sub> NO <sub>4</sub> S	d <sub>5</sub> -EtFOSAA	0.5	88±9%
4:2 fluorotelomer sulfonate	4:2 FTS	C <sub>6</sub> H <sub>5</sub> O <sub>3</sub> SF <sub>9</sub>	13C <sub>2</sub> -4:2 FTS	0.5	112±8%
6:2 fluorotelomer sulfonate	6:2 FTS	C <sub>6</sub> H <sub>5</sub> O <sub>3</sub> SF <sub>9</sub>	13C <sub>2</sub> -6:2 FTS	1	90±7%
8:2 fluorotelomer sulfonate	8:2 FTS	C <sub>8</sub> H <sub>5</sub> O <sub>3</sub> SF <sub>13</sub>	13C <sub>2</sub> -8:2 FTS	0.5	90±8%

Note: Analytical standards of LC target compounds were purchased from Wellington

Laboratories and product codes are grouped in different colors shown below.

PFAC-24PAR
MPFAC-24ES
*L-PFPrS
**L-PFOS
***br-PFOS



N-TMAmP-FHxSAE	N-TriMethylAmmin oPropyl- perFluoroHexylS ulfonAmido Ethanoic acid	C14H17F13N2O4S	C[N+](C)(C)CCC N(CC([O-])=O)S(=O)(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	Sulfonamide ; Carboxylic acid; Amine	6	√	√	
N-FBEAc	N-perFluoroButylEt hylAcetate	C8H7F9O2	CC(=O)OCCC(F)(F)C(F)(F)C(F)(F)C(F)(F)F	Ester	4	√		
N-EFHpSA-EMAC	N-EthylperFluoroHe ptylSulfonAmido- EthylMethACryla te	C15H14F15NO4S	CCN(CCOC(=O)C(C)=C)S(=O)(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	Sulfonamide ; Ester; Alkene	7	√	√	
N-FHxP-MAC	N-perFluoroHexylPr opanyl- MeACrylate	C13H11F13O2	CC(=C)C(=O)OCC CC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	Alkene; Ester	6	√		









**Table B6.3** Qualified hydrocarbon surfactants in AFFF using suspect-screening

Identification	Molecular formula	[M-H]-	Error (ppm)	Retention-Time	SMILES	Reference	Reference_DOI
C10-AE1S	C12H26O5S	281.14	0.59	13.54	CCCCCCCC COCCOS(=O)( O)=O		
C10-AE2S	C14H30O6S	325.17	0.82	14.32	OS(OCCOCCO CCCCCCCC C)(=O)=O		
C10-AE3S	C16H34O7S	369.19	-2.42	14.78	CCCCCCCC COCCOCCOC COS(=O)(=O)O		
C10-AE4S	C18H38O8S	413.22	-2.42	15.15	O=S(OCCOCC OCCOCCOCC CCCCCCCC)( O)=O		
C10-AE5S	C20H42O9S	457.25	-2.84	15.39	O=S(OCCOCC OCCOCCOCC OCCCCCCCC CC)(O)=O		
C10-AE6S	C22H46O10S	501.27	-3.72	15.61	O=S(OCCOCC OCCOCCOCC OCCOCCCCC CCCC)(O)=O		
C10-AE7S	C24H50O11S	545.30	-5.04	15.79	O=S(OCCOCC OCCOCCOCC OCCOCCOCC CCCCCCCC)( O)=O		
C12-AE1S	C14H30O5S	309.17	0.22	16.63	CCCCCCCC CCCOCCOS(= O)(=O)O	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-AE2S	C16H34O6S	353.20	-1.51	17.44	O=S(O)(=O)OC COCCOCCCC CCCCCCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374

C12-AE3S	C18H38O7S1	397.23	-1.63	17.81	O=S(O)(=O)OC COCCOCCOC CCCCCCCC CC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-AE4S	C20H42O8S	441.25	-2.57	18.26	O=S(O)(=O)OC COCCOCCOC COCCOCCOC CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-AE5S	C22H46O9S	485.28	-1.92	18.51	O=S(O)(=O)OC COCCOCCOC COCCOCCOC CCCCCCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-AE6S	C24H50O10S	529.31	-2.71	18.69	O=S(O)(=O)OC COCCOCCOC COCCOCCOC CCCCCCCC CC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-AE7S	C26H54O11S	573.33	-4.13	18.84	O=S(O)(=O)OC COCCOCCOC COCCOCCOC COCCOCCOC CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C10-LAS	C16H26O3S	297.15	-0.73	15.14	OS(=O)(=O)c1 ccc(cc1)C(CCC CC)CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C11-LAS	C17H28O3S	311.17	0.50	16.57	OS(=O)(=O)c1 ccc(cc1)C(CCC CCC)CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C12-LAS	C18H30O3S	325.18	-1.63	18.07	OS(=O)(=O)c1 ccc(cc1)C(CCC CCCC)CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374
C13-LAS	C19H32O3S	339.20	-0.67	19.45	OS(=O)(=O)c1 ccc(cc1)C(CCC CCCC)CCCC	Schymanski_et al_2014	dx.doi.org/10.1 021/es4044374

**Table B7.1** AFFF formulation comparison of total fluorine quantified in AFFF using HPLC-QTOF-MS and 19F NMR (Total fluorine quant)

AFFF(3M, 1999)-formulation analysis by LC-QTOF-MS and NMR			
Data File	AFFF formulation (g/L)	Std (g/L)	PFASs MW
PFPeA	0.11	0.00	263.98
L-PFBS	0.19	0.01	299.95
PFBA	0.09	0.00	213.99
PFHxA	0.11	0.02	313.98
L-PFPeS	0.18	0.05	349.95
PFHpA	0.08	0.00	363.98
PFHxSK	0.97	0.06	399.94
PFOA	0.11	0.00	413.97
L-PFHpS	0.15	0.01	449.94
PFNA	0.04	0.00	463.97
L-PFOS	3.09	0.25	499.94
br-PFOS	0.57	0.04	499.94
FOSA	0.09	0.00	498.95
L-PFNS	0.07	0.00	549.93
L-PFDS	0.05	0.00	599.93
Total Flourine determined by LC-MS (g/L)	3.78		
Total Flourine determined by NMR (g/L)	44.01		
Percent difference	91.4%		

**Table B7.2** Experimentally determined and molecular docking predicted HSA-PFAS association constants

Ligand	1E7G		1A06		Experimental result		
	GM Log Ka* (M <sup>-1</sup> )	GSD Log Ka** (M <sup>-1</sup> )	GM Log Ka* (M <sup>-1</sup> )	GSD Log Ka** (M <sup>-1</sup> )	log Ka (M <sup>-1</sup> )	SD log Ka (M <sup>-1</sup> )	Fitting model
PFBA	4.1	0.3	3.9	0.3	4.1	0.5	Langmuir isotherm
PFPeA	4.7	0.4	4.2	0.3	4.1	0.3	1st order association
PFHxA	5.1	0.4	4.5	0.4	4.6	0.2	1st order association
PFHpA	5.5	0.4	4.7	0.6	5.5	0.4	1st order association
PFOA	5.9	0.5	4.9	0.7	5.0	0.1	1st order association
PFNA	6.2	0.5	5.1	0.7	5.4	0.2	1st order association
PFDA	6.5	0.7	5.3	0.7	4.9	0.1	1st order association
PFUdA	6.7	0.6	5.3	0.9	4.6	0.2	1st order association
PFDoA	7.1	0.6	5.4	1.0	4.5	0.5	1st order association
PFTTrDA	7.3	0.7	5.3	0.9	4.3	0.5	1st order association
PFTeDA	7.5	0.7	5.4	1.0	4.0	0.9	1st order association
L-PFOS	6.3	0.5	5.0	0.7	4.7	0.5	Langmuir isotherm
PFBS	4.6	0.3	4.2	0.3	4.6	0.1	1st order association
PFPeS	5.0	0.4	4.4	0.4	4.8	0.5	1st order association
PFHxS	5.5	0.4	4.6	0.6	5.0	0.4	1st order association
PFHpS	5.9	0.4	4.8	0.7	4.6	0.3	1st order association
L-PFPrS	4.2	0.3	3.9	0.2	4.1	0.2	1st order association
PFNS	6.5	0.7	5.1	0.8	5.1	0.2	1st order association
PFDS	6.8	0.6	5.1	0.9	4.6	0.2	1st order association
FOSA	6.4	0.6	5.1	0.7	4.8	0.1	1st order association
MeFOSAA	6.0	0.4	4.7	0.7	5.1	0.2	1st order association
EtFOSAA	6.0	0.4	4.6	0.7	5.1	0.2	1st order association
4:2FTS	5.1	0.4	4.3	0.4	4.9	0.1	1st order association

6:2FTS	5.8	0.5	4.6	0.6	5.2	0.1	1st order association
8:2FTS	6.5	0.6	5.0	0.8	4.6	0.1	1st order association
HFPO-DA	5.3	0.4	4.5	0.6	NA	NA	NA
br-PFOS	NA	NA	NA	NA	4.5	0.3	Langmuir isotherm

Note:

1. Experimentally determined HFPO (GenX) binding affinity was not reported due to identification problem caused by in-source ionization. Further method optimization will be performed.
2. Experimentally determined br-PFOS was a mixture of branched PFOS. Since identification of individual compound was not possible, docking predicted association constant was not generated.

**Table B7.3** Kruskal-Wallis one-way pairwise ANOVA analysis on docking predicted HSA (1E7G) binding affinities of long-chain PFAS and hydrocarbon surfactants identified in AFFF (Kruskal-Wallis ANOVA)

Pairwise Comparisons of Long-chain Compounds identified in AFFF					
Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Significance	Adjusted Significance*
C11-LAS-C10-LAS	21.759	1048.391	0.021	0.983	1
C10-AE7S-C10-AE1S	92.604	842.478	0.11	0.912	1
PFUdA-PFDS	96.283	842.518	0.114	0.909	1
C12-LAS-C11-LAS	187.856	1015.934	0.185	0.853	1
C12-LAS-C10-LAS	209.615	1027.731	0.204	0.838	1
N-diMAmP-FHxSA-8:2 monoPAP-diEes	202.913	842.598	0.241	0.81	1
C12-AE2S-C12-AE4S	-246.269	842.478	-0.292	0.77	1
C10-AE3S-C10-AE5S	-314.585	842.478	-0.373	0.709	1
C12-AE1S-C12-AE5S	-360.259	842.478	-0.428	0.669	1
N-PFBS-MFPe-H-PFDS	492.219	842.637	0.584	0.559	1
C12-AE7S-C10-AE6S	975.727	842.478	1.158	0.247	1
C10-AE4S-C10-AE3S	1012.76	842.478	1.202	0.229	1
4-FHp-CycHxA-PFUdA	-1103.639	842.597	-1.31	0.19	1
4-FHp-CycHxA-PFDS	-1199.922	842.558	-1.424	0.154	1
C10-AE2S-C10-AE7S	-1228.741	842.478	-1.458	0.145	1
C10-AE2S-C10-AE1S	1321.345	842.478	1.568	0.117	1
C10-AE4S-C10-AE5S	-1327.345	842.478	-1.576	0.115	1
C10-AE5S-C12-AE7S	-1390.045	842.478	-1.65	0.099	1
C13-LAS-C12-LAS	1906.666	1002.543	1.902	0.057	1
C10-AE3S-C12-AE7S	-1704.63	842.478	-2.023	0.043	1
C13-LAS-C11-LAS	2094.522	1023.712	2.046	0.041	1
C13-LAS-C10-LAS	2116.281	1035.421	2.044	0.041	1
C10-LAS-N-diMAmP-FBSAP	-2028.462	957.341	-2.119	0.034	1
C12-AE4S-C12-AE1S	1818.028	842.478	2.158	0.031	1
C11-LAS-N-diMAmP-FBSAP	-2050.221	944.665	-2.17	0.03	1
H-PFDS-PFDA	-1903.655	842.478	-2.26	0.024	1
C10-AE6S-C10-AE2S	1914.03	842.478	2.272	0.023	1
C12-AE6S-C10-AE4S	1968.574	842.478	2.337	0.019	1
C12-LAS-N-diMAmP-FBSAP	-2238.077	921.683	-2.428	0.015	1
C12-AE3S-C12-AE2S	2077.825	842.478	2.466	0.014	1
C12-AE2S-C12-AE1S	2064.298	842.478	2.45	0.014	1
C12-AE4S-C12-AE5S	-2178.288	842.478	-2.586	0.01	1
C12-AE3S-C12-AE4S	-2324.094	842.478	-2.759	0.006	1
C10-AE5S-C10-AE6S	-2365.772	842.478	-2.808	0.005	1
N-PFBS-MFPe-PFDA	-2395.874	842.637	-2.843	0.004	1
C12-AE2S-C12-AE5S	-2424.557	842.478	-2.878	0.004	1
C10-AE3S-C10-AE6S	-2680.357	842.478	-3.182	0.001	0.637

C10-AE4S-C12-AE7S	-2717.39	842.478	-3.225	0.001	0.547
C12-AE7S-C10-AE2S	2889.757	842.478	3.43	0.001	0.262
C12-AE6S-C10-AE3S	2981.334	842.478	3.539	0	0.175
C10-AE6S-C10-AE7S	-3142.771	842.478	-3.73	0	0.083
C10-AE6S-C10-AE1S	3235.375	842.478	3.84	0	0.053
C12-AE6S-C10-AE5S	3295.919	842.478	3.912	0	0.04
C10-AE4S-C10-AE6S	-3693.117	842.478	-4.384	0	0.005
C13-LAS-N-diMAmP-FBSAP	-4144.743	930.249	-4.456	0	0.004
PFDaA-4-FHp-CycHxA	7618.954	842.558	9.043	0	0
PFDaA-PFUdA	-8722.593	842.518	-10.353	0	0
PFDaA-PFDS	8818.876	842.478	10.468	0	0
PFDaA-N-PFBS-MFPe	13778.626	842.637	16.352	0	0
PFDaA-H-PFDS	14270.845	842.478	16.939	0	0
PFDaA-PFDA	16174.5	842.478	19.199	0	0
PFDaA-N-FHxP-MAC	30458.303	842.478	36.153	0	0
PFDaA-N-diMAmP-FHxSA	38996.736	842.478	46.288	0	0
PFDaA-8:2 monoPAP-diEes	39199.65	842.598	46.522	0	0
PFDaA-Uridine-FB	-45707.209	937.182	-48.771	0	0
PFDaA-C13-LAS	56542.899	930.249	60.783	0	0
PFDaA-C12-LAS	58449.565	921.683	63.416	0	0
PFDaA-C11-LAS	58637.421	944.665	62.072	0	0
PFDaA-C10-LAS	58659.18	957.341	61.273	0	0
PFDaA-N-diMAmP-FBSAP	60687.642	842.478	72.035	0	0
PFDaA-C12-AE3S	91541.541	842.478	108.657	0	0
PFDaA-C12-AE2S	93619.366	842.478	111.124	0	0
PFDaA-C12-AE4S	93865.635	842.478	111.416	0	0
PFDaA-C12-AE1S	95683.663	842.478	113.574	0	0
PFDaA-C12-AE5S	96043.923	842.478	114.002	0	0
PFDaA-C12-AE6S	101006.818	842.478	119.893	0	0
PFDaA-C10-AE4S	102975.392	842.478	122.229	0	0
PFDaA-C10-AE3S	103988.152	842.478	123.431	0	0
PFDaA-C10-AE5S	104302.737	842.478	123.805	0	0
PFDaA-C12-AE7S	105692.782	842.478	125.455	0	0
PFDaA-C10-AE6S	106668.509	842.478	126.613	0	0
PFDaA-C10-AE2S	108582.539	842.478	128.885	0	0
PFDaA-C10-AE7S	109811.28	842.478	130.343	0	0
PFDaA-C10-AE1S	109903.884	842.478	130.453	0	0
4-FHp-CycHxA-N-PFBS-MFPe	-6159.672	842.717	-7.309	0	0
4-FHp-CycHxA-H-PFDS	-6651.891	842.558	-7.895	0	0
4-FHp-CycHxA-PFDA	-8555.546	842.558	-10.154	0	0
4-FHp-CycHxA-N-FHxP-MAC	-22839.349	842.558	-27.107	0	0
4-FHp-CycHxA-N-diMAmP-FHxSA	-31377.782	842.558	-37.241	0	0
4-FHp-CycHxA-8:2 monoPAP-diEes	-31580.695	842.677	-37.477	0	0

4-FHp-CycHxA-Uridine-FB	-38088.255	937.253	-40.638	0	0
4-FHp-CycHxA-C13-LAS	-48923.945	930.321	-52.588	0	0
4-FHp-CycHxA-C12-LAS	-50830.611	921.755	-55.145	0	0
4-FHp-CycHxA-C11-LAS	-51018.467	944.736	-54.003	0	0
4-FHp-CycHxA-C10-LAS	-51040.226	957.411	-53.311	0	0
4-FHp-CycHxA-N-diMAMP-FBSAP	-53068.688	842.558	-62.985	0	0
4-FHp-CycHxA-C12-AE3S	-83922.587	842.558	-99.605	0	0
4-FHp-CycHxA-C12-AE2S	-86000.411	842.558	-102.071	0	0
4-FHp-CycHxA-C12-AE4S	-86246.681	842.558	-102.363	0	0
4-FHp-CycHxA-C12-AE1S	-88064.709	842.558	-104.521	0	0
4-FHp-CycHxA-C12-AE5S	-88424.969	842.558	-104.948	0	0
4-FHp-CycHxA-C12-AE6S	-93387.864	842.558	-110.839	0	0
4-FHp-CycHxA-C10-AE4S	-95356.437	842.558	-113.175	0	0
4-FHp-CycHxA-C10-AE3S	-96369.197	842.558	-114.377	0	0
4-FHp-CycHxA-C10-AE5S	-96683.782	842.558	-114.75	0	0
4-FHp-CycHxA-C12-AE7S	-98073.828	842.558	-116.4	0	0
4-FHp-CycHxA-C10-AE6S	-99049.555	842.558	-117.558	0	0
4-FHp-CycHxA-C10-AE2S	-100963.585	842.558	-119.83	0	0
4-FHp-CycHxA-C10-AE7S	-102192.326	842.558	-121.288	0	0
4-FHp-CycHxA-C10-AE1S	-102284.929	842.558	-121.398	0	0
PFUdA-N-PFBS-MFPe	5056.033	842.677	6	0	0
PFUdA-H-PFDS	5548.252	842.518	6.585	0	0
PFUdA-PFDA	7451.907	842.518	8.845	0	0
PFUdA-N-FHxP-MAC	21735.71	842.518	25.799	0	0
PFUdA-N-diMAMP-FHxSA	30274.144	842.518	35.933	0	0
PFUdA-8:2 monoPAP-diEes	30477.057	842.637	36.169	0	0
PFUdA-Uridine-FB	-36984.616	937.218	-39.462	0	0
PFUdA-C13-LAS	47820.306	930.285	51.404	0	0
PFUdA-C12-LAS	49726.972	921.719	53.95	0	0
PFUdA-C11-LAS	49914.828	944.7	52.837	0	0
PFUdA-C10-LAS	49936.587	957.376	52.16	0	0
PFUdA-N-diMAMP-FBSAP	51965.05	842.518	61.678	0	0

PFUdA-C12-AE3S	82818.948	842.518	98.299	0	0
PFUdA-C12-AE2S	84896.773	842.518	100.766	0	0
PFUdA-C12-AE4S	85143.042	842.518	101.058	0	0
PFUdA-C12-AE1S	86961.071	842.518	103.216	0	0
PFUdA-C12-AE5S	87321.33	842.518	103.643	0	0
PFUdA-C12-AE6S	92284.225	842.518	109.534	0	0
PFUdA-C10-AE4S	94252.799	842.518	111.87	0	0
PFUdA-C10-AE3S	95265.559	842.518	113.072	0	0
PFUdA-C10-AE5S	95580.144	842.518	113.446	0	0
PFUdA-C12-AE7S	96970.189	842.518	115.096	0	0
PFUdA-C10-AE6S	97945.916	842.518	116.254	0	0
PFUdA-C10-AE2S	99859.946	842.518	118.526	0	0
PFUdA-C10-AE7S	101088.687	842.518	119.984	0	0
PFUdA-C10-AE1S	101181.291	842.518	120.094	0	0
PFDS-N-PFBS-MFPe	4959.75	842.637	5.886	0	0
PFDS-H-PFDS	5451.969	842.478	6.471	0	0
PFDS-PFDA	7355.624	842.478	8.731	0	0
PFDS-N-FHxP-MAC	21639.427	842.478	25.685	0	0
PFDS-N-diMAmP-FHxSA	30177.861	842.478	35.82	0	0
PFDS-8:2 monoPAP-diEes	30380.774	842.598	36.056	0	0
PFDS-Uridine-FB	-36888.333	937.182	-39.361	0	0
PFDS-C13-LAS	47724.023	930.249	51.302	0	0
PFDS-C12-LAS	49630.689	921.683	53.848	0	0
PFDS-C11-LAS	49818.545	944.665	52.737	0	0
PFDS-C10-LAS	49840.304	957.341	52.061	0	0
PFDS-N-diMAmP-FBSAP	51868.767	842.478	61.567	0	0
PFDS-C12-AE3S	82722.665	842.478	98.19	0	0
PFDS-C12-AE2S	84800.49	842.478	100.656	0	0
PFDS-C12-AE4S	85046.759	842.478	100.948	0	0
PFDS-C12-AE1S	86864.788	842.478	103.106	0	0
PFDS-C12-AE5S	87225.047	842.478	103.534	0	0
PFDS-C12-AE6S	92187.942	842.478	109.425	0	0
PFDS-C10-AE4S	94156.516	842.478	111.761	0	0
PFDS-C10-AE3S	95169.276	842.478	112.964	0	0
PFDS-C10-AE5S	95483.861	842.478	113.337	0	0
PFDS-C12-AE7S	96873.906	842.478	114.987	0	0
PFDS-C10-AE6S	97849.633	842.478	116.145	0	0
PFDS-C10-AE2S	99763.663	842.478	118.417	0	0
PFDS-C10-AE7S	100992.404	842.478	119.875	0	0
PFDS-C10-AE1S	101085.008	842.478	119.985	0	0
N-PFBS-MFPe-N-FHxP-MAC	16679.677	842.637	19.795	0	0
N-PFBS-MFPe-N-diMAmP-FHxSA	-25218.11	842.637	-29.928	0	0
N-PFBS-MFPe-8:2 monoPAP-diEes	25421.023	842.757	30.164	0	0
N-PFBS-MFPe-Uridine-FB	-31928.583	937.325	-34.064	0	0

N-PFBS-MFPe-C13-LAS	42764.273	930.393	45.964	0	0
N-PFBS-MFPe-C12-LAS	44670.939	921.828	48.459	0	0
N-PFBS-MFPe-C11-LAS	44858.795	944.807	47.479	0	0
N-PFBS-MFPe-C10-LAS	44880.554	957.481	46.874	0	0
N-PFBS-MFPe-N-diMAmP-FBSAP	-46909.016	842.637	-55.669	0	0
N-PFBS-MFPe-C12-AE3S	77762.915	842.637	92.285	0	0
N-PFBS-MFPe-C12-AE2S	79840.739	842.637	94.751	0	0
N-PFBS-MFPe-C12-AE4S	80087.009	842.637	95.043	0	0
N-PFBS-MFPe-C12-AE1S	81905.037	842.637	97.201	0	0
N-PFBS-MFPe-C12-AE5S	82265.297	842.637	97.628	0	0
N-PFBS-MFPe-C12-AE6S	87228.192	842.637	103.518	0	0
N-PFBS-MFPe-C10-AE4S	89196.765	842.637	105.854	0	0
N-PFBS-MFPe-C10-AE3S	90209.525	842.637	107.056	0	0
N-PFBS-MFPe-C10-AE5S	90524.11	842.637	107.429	0	0
N-PFBS-MFPe-C12-AE7S	91914.156	842.637	109.079	0	0
N-PFBS-MFPe-C10-AE6S	92889.882	842.637	110.237	0	0
N-PFBS-MFPe-C10-AE2S	94803.913	842.637	112.509	0	0
N-PFBS-MFPe-C10-AE7S	96032.654	842.637	113.967	0	0
N-PFBS-MFPe-C10-AE1S	96125.257	842.637	114.077	0	0
H-PFDS-N-FHxP-MAC	-16187.458	842.478	-19.214	0	0
H-PFDS-N-diMAmP-FHxSA	-24725.891	842.478	-29.349	0	0
H-PFDS-8:2 monoPAP-diEes	24928.804	842.598	29.586	0	0
H-PFDS-Uridine-FB	-31436.364	937.182	-33.544	0	0
H-PFDS-C13-LAS	42272.054	930.249	45.442	0	0
H-PFDS-C12-LAS	44178.72	921.683	47.933	0	0
H-PFDS-C11-LAS	44366.576	944.665	46.965	0	0
H-PFDS-C10-LAS	44388.335	957.341	46.366	0	0
H-PFDS-N-diMAmP-FBSAP	-46416.797	842.478	-55.096	0	0
H-PFDS-C12-AE3S	77270.696	842.478	91.718	0	0
H-PFDS-C12-AE2S	79348.521	842.478	94.185	0	0
H-PFDS-C12-AE4S	79594.79	842.478	94.477	0	0
H-PFDS-C12-AE1S	81412.818	842.478	96.635	0	0
H-PFDS-C12-AE5S	81773.078	842.478	97.063	0	0

H-PFDS-C12-AE6S	86735.973	842.478	102.953	0	0
H-PFDS-C10-AE4S	88704.547	842.478	105.29	0	0
H-PFDS-C10-AE3S	89717.307	842.478	106.492	0	0
H-PFDS-C10-AE5S	90031.892	842.478	106.866	0	0
H-PFDS-C12-AE7S	91421.937	842.478	108.516	0	0
H-PFDS-C10-AE6S	92397.664	842.478	109.674	0	0
H-PFDS-C10-AE2S	94311.694	842.478	111.946	0	0
H-PFDS-C10-AE7S	95540.435	842.478	113.404	0	0
H-PFDS-C10-AE1S	95633.039	842.478	113.514	0	0
PFDA-N-FHxP-MAC	14283.803	842.478	16.955	0	0
PFDA-N-diMAmP-FHxSA	22822.236	842.478	27.089	0	0
PFDA-8:2 monoPAP-diEes	23025.149	842.598	27.326	0	0
PFDA-Uridine-FB	-29532.709	937.182	-31.512	0	0
PFDA-C13-LAS	40368.399	930.249	43.395	0	0
PFDA-C12-LAS	42275.065	921.683	45.867	0	0
PFDA-C11-LAS	42462.921	944.665	44.95	0	0
PFDA-C10-LAS	42484.68	957.341	44.378	0	0
PFDA-N-diMAmP-FBSAP	44513.142	842.478	52.836	0	0
PFDA-C12-AE3S	75367.041	842.478	89.459	0	0
PFDA-C12-AE2S	77444.865	842.478	91.925	0	0
PFDA-C12-AE4S	77691.135	842.478	92.217	0	0
PFDA-C12-AE1S	79509.163	842.478	94.375	0	0
PFDA-C12-AE5S	79869.423	842.478	94.803	0	0
PFDA-C12-AE6S	84832.318	842.478	100.694	0	0
PFDA-C10-AE4S	86800.891	842.478	103.03	0	0
PFDA-C10-AE3S	87813.651	842.478	104.233	0	0
PFDA-C10-AE5S	88128.236	842.478	104.606	0	0
PFDA-C12-AE7S	89518.282	842.478	106.256	0	0
PFDA-C10-AE6S	90494.009	842.478	107.414	0	0
PFDA-C10-AE2S	92408.039	842.478	109.686	0	0
PFDA-C10-AE7S	93636.78	842.478	111.144	0	0
PFDA-C10-AE1S	93729.383	842.478	111.254	0	0
N-FHxP-MAC-N-diMAmP-FHxSA	-8538.433	842.478	-10.135	0	0
N-FHxP-MAC-8:2 monoPAP-diEes	8741.346	842.598	10.374	0	0
N-FHxP-MAC-Uridine-FB	-15248.906	937.182	-16.271	0	0
N-FHxP-MAC-C13-LAS	26084.596	930.249	28.04	0	0
N-FHxP-MAC-C12-LAS	27991.262	921.683	30.37	0	0
N-FHxP-MAC-C11-LAS	28179.118	944.665	29.83	0	0
N-FHxP-MAC-C10-LAS	28200.877	957.341	29.457	0	0
N-FHxP-MAC-N-diMAmP-FBSAP	-30229.339	842.478	-35.881	0	0
N-FHxP-MAC-C12-AE3S	61083.238	842.478	72.504	0	0
N-FHxP-MAC-C12-AE2S	63161.063	842.478	74.971	0	0
N-FHxP-MAC-C12-AE4S	63407.332	842.478	75.263	0	0
N-FHxP-MAC-C12-AE1S	65225.36	842.478	77.421	0	0

N-FHxP-MAC-C12-AE5S	65585.62	842.478	77.848	0	0
N-FHxP-MAC-C12-AE6S	70548.515	842.478	83.739	0	0
N-FHxP-MAC-C10-AE4S	72517.089	842.478	86.076	0	0
N-FHxP-MAC-C10-AE3S	73529.848	842.478	87.278	0	0
N-FHxP-MAC-C10-AE5S	73844.434	842.478	87.651	0	0
N-FHxP-MAC-C12-AE7S	75234.479	842.478	89.301	0	0
N-FHxP-MAC-C10-AE6S	76210.206	842.478	90.46	0	0
N-FHxP-MAC-C10-AE2S	78124.236	842.478	92.731	0	0
N-FHxP-MAC-C10-AE7S	79352.977	842.478	94.19	0	0
N-FHxP-MAC-C10-AE1S	79445.58	842.478	94.3	0	0
N-diMAmP-FHxSA-Uridine-FB	-6710.472	937.182	-7.16	0	0
N-diMAmP-FHxSA-C13-LAS	17546.163	930.249	18.862	0	0
N-diMAmP-FHxSA-C12-LAS	19452.829	921.683	21.106	0	0
N-diMAmP-FHxSA-C11-LAS	19640.685	944.665	20.791	0	0
N-diMAmP-FHxSA-C10-LAS	19662.444	957.341	20.539	0	0
N-diMAmP-FHxSA-N-diMAmP-FBSAP	21690.906	842.478	25.747	0	0
N-diMAmP-FHxSA-C12-AE3S	52544.805	842.478	62.369	0	0
N-diMAmP-FHxSA-C12-AE2S	54622.629	842.478	64.836	0	0
N-diMAmP-FHxSA-C12-AE4S	54868.899	842.478	65.128	0	0
N-diMAmP-FHxSA-C12-AE1S	56686.927	842.478	67.286	0	0
N-diMAmP-FHxSA-C12-AE5S	57047.186	842.478	67.714	0	0
N-diMAmP-FHxSA-C12-AE6S	62010.082	842.478	73.604	0	0
N-diMAmP-FHxSA-C10-AE4S	63978.655	842.478	75.941	0	0
N-diMAmP-FHxSA-C10-AE3S	64991.415	842.478	77.143	0	0
N-diMAmP-FHxSA-C10-AE5S	65306	842.478	77.517	0	0
N-diMAmP-FHxSA-C12-AE7S	66696.045	842.478	79.167	0	0
N-diMAmP-FHxSA-C10-AE6S	67671.772	842.478	80.325	0	0
N-diMAmP-FHxSA-C10-AE2S	69585.803	842.478	82.597	0	0
N-diMAmP-FHxSA-C10-AE7S	70814.544	842.478	84.055	0	0
N-diMAmP-FHxSA-C10-AE1S	70907.147	842.478	84.165	0	0

8:2 monoPAP-diEes-Uridine-FB	-6507.559	937.289	-6.943	0	0
8:2 monoPAP-diEes-C13-LAS	-17343.25	930.357	-18.641	0	0
8:2 monoPAP-diEes-C12-LAS	-19249.916	921.792	-20.883	0	0
8:2 monoPAP-diEes-C11-LAS	-19437.772	944.771	-20.574	0	0
8:2 monoPAP-diEes-C10-LAS	-19459.531	957.446	-20.324	0	0
8:2 monoPAP-diEes-N-diMAmP-FBSAP	-21487.993	842.598	-25.502	0	0
8:2 monoPAP-diEes-C12-AE3S	-52341.891	842.598	-62.12	0	0
8:2 monoPAP-diEes-C12-AE2S	-54419.716	842.598	-64.586	0	0
8:2 monoPAP-diEes-C12-AE4S	-54665.985	842.598	-64.878	0	0
8:2 monoPAP-diEes-C12-AE1S	-56484.014	842.598	-67.036	0	0
8:2 monoPAP-diEes-C12-AE5S	-56844.273	842.598	-67.463	0	0
8:2 monoPAP-diEes-C12-AE6S	-61807.169	842.598	-73.353	0	0
8:2 monoPAP-diEes-C10-AE4S	-63775.742	842.598	-75.689	0	0
8:2 monoPAP-diEes-C10-AE3S	-64788.502	842.598	-76.891	0	0
8:2 monoPAP-diEes-C10-AE5S	-65103.087	842.598	-77.265	0	0
8:2 monoPAP-diEes-C12-AE7S	-66493.132	842.598	-78.914	0	0
8:2 monoPAP-diEes-C10-AE6S	-67468.859	842.598	-80.072	0	0
8:2 monoPAP-diEes-C10-AE2S	-69382.89	842.598	-82.344	0	0
8:2 monoPAP-diEes-C10-AE7S	-70611.631	842.598	-83.802	0	0
8:2 monoPAP-diEes-C10-AE1S	-70704.234	842.598	-83.912	0	0
Uridine-FB-C13-LAS	10835.69	1016.81	10.657	0	0
Uridine-FB-C12-LAS	12742.356	1008.979	12.629	0	0
Uridine-FB-C11-LAS	12930.212	1030.016	12.553	0	0
Uridine-FB-C10-LAS	12951.971	1041.654	12.434	0	0
Uridine-FB-N-diMAmP-FBSAP	14980.434	937.182	15.985	0	0
Uridine-FB-C12-AE3S	45834.332	937.182	48.907	0	0
Uridine-FB-C12-AE2S	47912.157	937.182	51.124	0	0
Uridine-FB-C12-AE4S	48158.426	937.182	51.386	0	0
Uridine-FB-C12-AE1S	49976.455	937.182	53.326	0	0
Uridine-FB-C12-AE5S	50336.714	937.182	53.711	0	0

Uridine-FB-C12-AE6S	55299.609	937.182	59.006	0	0
Uridine-FB-C10-AE4S	57268.183	937.182	61.107	0	0
Uridine-FB-C10-AE3S	58280.943	937.182	62.187	0	0
Uridine-FB-C10-AE5S	58595.528	937.182	62.523	0	0
Uridine-FB-C12-AE7S	59985.573	937.182	64.006	0	0
Uridine-FB-C10-AE6S	60961.3	937.182	65.047	0	0
Uridine-FB-C10-AE2S	62875.33	937.182	67.09	0	0
Uridine-FB-C10-AE7S	64104.071	937.182	68.401	0	0
Uridine-FB-C10-AE1S	64196.675	937.182	68.5	0	0
C13-LAS-C12-AE3S	34998.642	930.249	37.623	0	0
C13-LAS-C12-AE2S	37076.467	930.249	39.857	0	0
C13-LAS-C12-AE4S	37322.736	930.249	40.121	0	0
C13-LAS-C12-AE1S	39140.764	930.249	42.076	0	0
C13-LAS-C12-AE5S	39501.024	930.249	42.463	0	0
C13-LAS-C12-AE6S	44463.919	930.249	47.798	0	0
C13-LAS-C10-AE4S	46432.493	930.249	49.914	0	0
C13-LAS-C10-AE3S	47445.252	930.249	51.003	0	0
C13-LAS-C10-AE5S	47759.838	930.249	51.341	0	0
C13-LAS-C12-AE7S	49149.883	930.249	52.835	0	0
C13-LAS-C10-AE6S	50125.61	930.249	53.884	0	0
C13-LAS-C10-AE2S	52039.64	930.249	55.942	0	0
C13-LAS-C10-AE7S	53268.381	930.249	57.263	0	0
C13-LAS-C10-AE1S	53360.985	930.249	57.362	0	0
C12-LAS-C12-AE3S	33091.976	921.683	35.904	0	0
C12-LAS-C12-AE2S	35169.8	921.683	38.158	0	0
C12-LAS-C12-AE4S	35416.07	921.683	38.425	0	0
C12-LAS-C12-AE1S	37234.098	921.683	40.398	0	0
C12-LAS-C12-AE5S	37594.358	921.683	40.789	0	0
C12-LAS-C12-AE6S	42557.253	921.683	46.173	0	0
C12-LAS-C10-AE4S	44525.826	921.683	48.309	0	0
C12-LAS-C10-AE3S	45538.586	921.683	49.408	0	0
C12-LAS-C10-AE5S	45853.172	921.683	49.749	0	0
C12-LAS-C12-AE7S	47243.217	921.683	51.258	0	0
C12-LAS-C10-AE6S	48218.944	921.683	52.316	0	0
C12-LAS-C10-AE2S	50132.974	921.683	54.393	0	0
C12-LAS-C10-AE7S	51361.715	921.683	55.726	0	0
C12-LAS-C10-AE1S	51454.318	921.683	55.827	0	0
C11-LAS-C12-AE3S	-32904.12	944.665	-34.832	0	0
C11-LAS-C12-AE2S	-34981.945	944.665	-37.031	0	0
C11-LAS-C12-AE4S	-35228.214	944.665	-37.292	0	0
C11-LAS-C12-AE1S	-37046.242	944.665	-39.216	0	0
C11-LAS-C12-AE5S	-37406.502	944.665	-39.598	0	0
C11-LAS-C12-AE6S	-42369.397	944.665	-44.851	0	0
C11-LAS-C10-AE4S	44337.971	944.665	46.935	0	0
C11-LAS-C10-AE3S	45350.73	944.665	48.007	0	0
C11-LAS-C10-AE5S	45665.316	944.665	48.34	0	0
C11-LAS-C12-AE7S	-47055.361	944.665	-49.812	0	0
C11-LAS-C10-AE6S	48031.088	944.665	50.845	0	0
C11-LAS-C10-AE2S	49945.118	944.665	52.871	0	0
C11-LAS-C10-AE7S	51173.859	944.665	54.171	0	0

C11-LAS-C10-AE1S	51266.463	944.665	54.269	0	0
C10-LAS-C12-AE3S	-32882.361	957.341	-34.348	0	0
C10-LAS-C12-AE2S	-34960.186	957.341	-36.518	0	0
C10-LAS-C12-AE4S	-35206.455	957.341	-36.775	0	0
C10-LAS-C12-AE1S	-37024.483	957.341	-38.674	0	0
C10-LAS-C12-AE5S	-37384.743	957.341	-39.051	0	0
C10-LAS-C12-AE6S	-42347.638	957.341	-44.235	0	0
C10-LAS-C10-AE4S	44316.211	957.341	46.291	0	0
C10-LAS-C10-AE3S	45328.971	957.341	47.349	0	0
C10-LAS-C10-AE5S	45643.557	957.341	47.677	0	0
C10-LAS-C12-AE7S	-47033.602	957.341	-49.129	0	0
C10-LAS-C10-AE6S	48009.329	957.341	50.149	0	0
C10-LAS-C10-AE2S	49923.359	957.341	52.148	0	0
C10-LAS-C10-AE7S	51152.1	957.341	53.431	0	0
C10-LAS-C10-AE1S	51244.703	957.341	53.528	0	0
N-diMAmP-FBSAP-C12-AE3S	30853.899	842.478	36.623	0	0
N-diMAmP-FBSAP-C12-AE2S	32931.723	842.478	39.089	0	0
N-diMAmP-FBSAP-C12-AE4S	33177.992	842.478	39.381	0	0
N-diMAmP-FBSAP-C12-AE1S	34996.021	842.478	41.539	0	0
N-diMAmP-FBSAP-C12-AE5S	35356.28	842.478	41.967	0	0
N-diMAmP-FBSAP-C12-AE6S	40319.176	842.478	47.858	0	0
N-diMAmP-FBSAP-C10-AE4S	42287.749	842.478	50.194	0	0
N-diMAmP-FBSAP-C10-AE3S	43300.509	842.478	51.397	0	0
N-diMAmP-FBSAP-C10-AE5S	43615.094	842.478	51.77	0	0
N-diMAmP-FBSAP-C12-AE7S	45005.139	842.478	53.42	0	0
N-diMAmP-FBSAP-C10-AE6S	45980.866	842.478	54.578	0	0
N-diMAmP-FBSAP-C10-AE2S	47894.897	842.478	56.85	0	0
N-diMAmP-FBSAP-C10-AE7S	49123.638	842.478	58.309	0	0
N-diMAmP-FBSAP-C10-AE1S	49216.241	842.478	58.418	0	0
C12-AE3S-C12-AE1S	4142.122	842.478	4.917	0	0
C12-AE3S-C12-AE5S	-4502.382	842.478	-5.344	0	0
C12-AE3S-C12-AE6S	-9465.277	842.478	-11.235	0	0
C12-AE3S-C10-AE4S	11433.851	842.478	13.572	0	0
C12-AE3S-C10-AE3S	12446.611	842.478	14.774	0	0
C12-AE3S-C10-AE5S	12761.196	842.478	15.147	0	0
C12-AE3S-C12-AE7S	-14151.241	842.478	-16.797	0	0
C12-AE3S-C10-AE6S	15126.968	842.478	17.955	0	0

C12-AE3S-C10-AE2S	17040.998	842.478	20.227	0	0
C12-AE3S-C10-AE7S	18269.739	842.478	21.686	0	0
C12-AE3S-C10-AE1S	18362.343	842.478	21.796	0	0
C12-AE2S-C12-AE6S	-7387.452	842.478	-8.769	0	0
C12-AE2S-C10-AE4S	9356.026	842.478	11.105	0	0
C12-AE2S-C10-AE3S	10368.786	842.478	12.307	0	0
C12-AE2S-C10-AE5S	10683.371	842.478	12.681	0	0
C12-AE2S-C12-AE7S	-12073.416	842.478	-14.331	0	0
C12-AE2S-C10-AE6S	13049.143	842.478	15.489	0	0
C12-AE2S-C10-AE2S	14963.173	842.478	17.761	0	0
C12-AE2S-C10-AE7S	16191.914	842.478	19.219	0	0
C12-AE2S-C10-AE1S	16284.518	842.478	19.329	0	0
C12-AE4S-C12-AE6S	-7141.183	842.478	-8.476	0	0
C12-AE4S-C10-AE4S	9109.757	842.478	10.813	0	0
C12-AE4S-C10-AE3S	10122.517	842.478	12.015	0	0
C12-AE4S-C10-AE5S	10437.102	842.478	12.389	0	0
C12-AE4S-C12-AE7S	-11827.147	842.478	-14.039	0	0
C12-AE4S-C10-AE6S	12802.874	842.478	15.197	0	0
C12-AE4S-C10-AE2S	14716.904	842.478	17.469	0	0
C12-AE4S-C10-AE7S	15945.645	842.478	18.927	0	0
C12-AE4S-C10-AE1S	16038.249	842.478	19.037	0	0
C12-AE1S-C12-AE6S	-5323.155	842.478	-6.318	0	0
C12-AE1S-C10-AE4S	7291.728	842.478	8.655	0	0
C12-AE1S-C10-AE3S	8304.488	842.478	9.857	0	0
C12-AE1S-C10-AE5S	8619.073	842.478	10.231	0	0
C12-AE1S-C12-AE7S	-10009.118	842.478	-11.881	0	0
C12-AE1S-C10-AE6S	10984.845	842.478	13.039	0	0
C12-AE1S-C10-AE2S	12898.876	842.478	15.311	0	0
C12-AE1S-C10-AE7S	14127.617	842.478	16.769	0	0
C12-AE1S-C10-AE1S	14220.22	842.478	16.879	0	0
C12-AE5S-C12-AE6S	-4962.895	842.478	-5.891	0	0
C12-AE5S-C10-AE4S	6931.469	842.478	8.227	0	0
C12-AE5S-C10-AE3S	7944.229	842.478	9.43	0	0
C12-AE5S-C10-AE5S	8258.814	842.478	9.803	0	0
C12-AE5S-C12-AE7S	-9648.859	842.478	-11.453	0	0
C12-AE5S-C10-AE6S	10624.586	842.478	12.611	0	0
C12-AE5S-C10-AE2S	12538.616	842.478	14.883	0	0
C12-AE5S-C10-AE7S	13767.357	842.478	16.342	0	0
C12-AE5S-C10-AE1S	13859.961	842.478	16.451	0	0
C12-AE6S-C12-AE7S	-4685.964	842.478	-5.562	0	0
C12-AE6S-C10-AE6S	5661.691	842.478	6.72	0	0
C12-AE6S-C10-AE2S	7575.721	842.478	8.992	0	0
C12-AE6S-C10-AE7S	8804.462	842.478	10.451	0	0
C12-AE6S-C10-AE1S	8897.066	842.478	10.561	0	0
C10-AE4S-C10-AE2S	5607.147	842.478	6.656	0	0
C10-AE4S-C10-AE7S	-6835.888	842.478	-8.114	0	0
C10-AE4S-C10-AE1S	6928.492	842.478	8.224	0	0
C10-AE3S-C10-AE2S	4594.388	842.478	5.453	0	0
C10-AE3S-C10-AE7S	-5823.128	842.478	-6.912	0	0
C10-AE3S-C10-AE1S	5915.732	842.478	7.022	0	0

C10-AE5S-C10-AE2S	4279.802	842.478	5.08	0	0
C10-AE5S-C10-AE7S	-5508.543	842.478	-6.539	0	0
C10-AE5S-C10-AE1S	5601.147	842.478	6.648	0	0
C12-AE7S-C10-AE7S	4118.498	842.478	4.889	0	0
C12-AE7S-C10-AE1S	4211.102	842.478	4.998	0	0

<b>Hypothesis Test Summary</b>				
	Null Hypothesis	Test	Sig. <sup>a,b</sup>	Decision
1	The distribution of Affinities is the same across categories of Compounds.	Independent-Samples Kruskal-Wallis Test	0.000	Reject the null hypothesis.
a. The significance level is .050.				
b. Asymptotic significance is displayed.				

<b>Independent-Samples Kruskal-Wallis Test Summary</b>	
Total N	156192
Test Statistic	123241.926 <sup>a</sup>
Degree Of Freedom	29
Asymptotic Sig. (2-sided test)	0.000
a. The test statistic is adjusted for ties.	

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .050. \*Significance values have been adjusted by the Bonferroni correction for multiple tests. Colored cells represent the experimentally determined bound vs unbound compounds that failed to show significantly different docking predicted binding affinities.

## References

- Ackley, D. C., Rockich, K. T., & Baker, T. R. (2004). Optimazation of Drug Discovery: In Vitro Methods. In *Metabolic Stability Assessed by Liver Microsomes and Hepatocytes*.
- Allendorf, F., Berger, U., Goss, K. U., & Ulrich, N. (2019). Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. *Environmental Science: Processes and Impacts*, 21(11), 1852–1863. <https://doi.org/10.1039/c9em00290a>
- 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. (2017). Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater. *Environmental Science and Technology*, 51(4), 2047–2057. <https://doi.org/10.1021/acs.est.6b05843>
- Curry, S. (2003). Plasma albumin as a fatty acid carrier. *Advances in Molecular and Cell Biology*, 33, 29–46.
- Gunasekar, S., Lee, J., Soudry, D., & Srebro, N. (2018). Characterizing implicit bias in terms of optimization geometry. *35th International Conference on Machine Learning, ICML 2018*, 4, 2932–2955.
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Advances in Mathematics*, 4(17). <https://doi.org/10.1016/j.aim.2014.05.019>
- Hirs, C. H. W., Stein, W. H., & Moore, S. (1954). The Free Amino Acids of Human Blood Plasma. *Journal of Biological Chemistry*, 211, 941–950.
- Luo, Z., Shi, X., Hu, Q., Zhao, B., & Huang, M. (2012). Structural evidence of perfluorooctane sulfonate transport by human serum albumin. *Chemical Research in Toxicology*, 25(5), 990–992. <https://doi.org/10.1021/tx300112p>
- Moschet, C., Anumol, T., Lew, B. M., Bennett, D. H., & Young, T. M. (2018). Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. *Environmental Science and Technology*, 52(5), 2878–2887. <https://doi.org/10.1021/acs.est.7b05767>
- Muñoz, A., Kral, R., & Schimmel, H. (2011). Quantification of protein calibrants by amino acid analysis using isotope dilution mass spectrometry. *Analytical Biochemistry*, 408(1), 124–131. <https://doi.org/10.1016/j.ab.2010.08.037>
- Mustăţea, G., Ungureanu, E. L., & Iorga, E. (2019). Protein acidic hydrolysis for amino acids analysis in food - progress over time: A short review. *Journal of Hygienic Engineering and Design*, 26(April), 81–87.
- Ng, C. A., & Hungerbuehler, K. (2015). Exploring the Use of Molecular Docking to Identify Bioaccumulative Perfluorinated Alkyl Acids (PFAAs). *Environmental Science and Technology*, 49(20), 12306–12314. <https://doi.org/10.1021/acs.est.5b03000>
- Otter, D. E. (2012). Standardised methods for amino acid analysis of food. *British Journal of Nutrition*, 108(SUPPL. 2), 230–237. <https://doi.org/10.1017/S0007114512002486>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>

- Sander, T., Freyss, J., von Korff, M., & Rufener, C. (2015). DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling*, 55(2), 460–473. <https://doi.org/10.1021/ci500588j>
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology*, 48(4), 2097–2098. <https://doi.org/10.1021/es5002105>
- Snow, A. W., Giles, S., Hinnant, K. M., Farley, J. P., & Ananth, R. (2017). *Quantification of Fluorine Content in AFFF Concentrates*.
- vanden Heuvel, J. P., Kuslikis, B. I., & Peterson, R. E. (1992). Covalent Binding of Perflourinated Fatty Acids to Proteins in the Plasma, Liver and Testes of Rats. *Chemical-Biological Interactions*, 82, 317–328.
- Wahl, J., Freyss, J., von Korff, M., & Sander, T. (2019). Accuracy evaluation and addition of improved dihedral parameters for the MMFF94s. *Journal of Cheminformatics*, 11(1), 1–10. <https://doi.org/10.1186/s13321-019-0371-6>

## APPENDIX C: SUPPLEMENTARY INFORMATION FOR CHAPTER 4

### Table of content

#### Figures

Figure C1. Receiver Operating Characteristic (ROC) curve of k-Nearest Neighbourhood (kNN)

Figure C2. Receiver Operating Characteristic (ROC) curve of Support Vector Classifier (SVC)

Figure C3. Receiver Operating Characteristic (ROC) curve of Random Forests (RF)

Figure C4. Receiver Operating Characteristic (ROC) curve of Bernoulli Naïve Bayes classifier (NB)

Figure C5. Receiver Operating Characteristic (ROC) curve of Gradient Boosting (GB)

Figure C6. Receiver Operating Characteristic (ROC) curve of Multilayer Perceptron (MLP) neural networks

Figure C7. Predicting performance of training and validation datasets of different testing models

#### Tables

Table C1. Summary of the hyperparameters for each machine learning model

Table C2. BioTox prediction for manufactured and used OECD PFAS in PubChem within the application domain and over 95% probability for classification

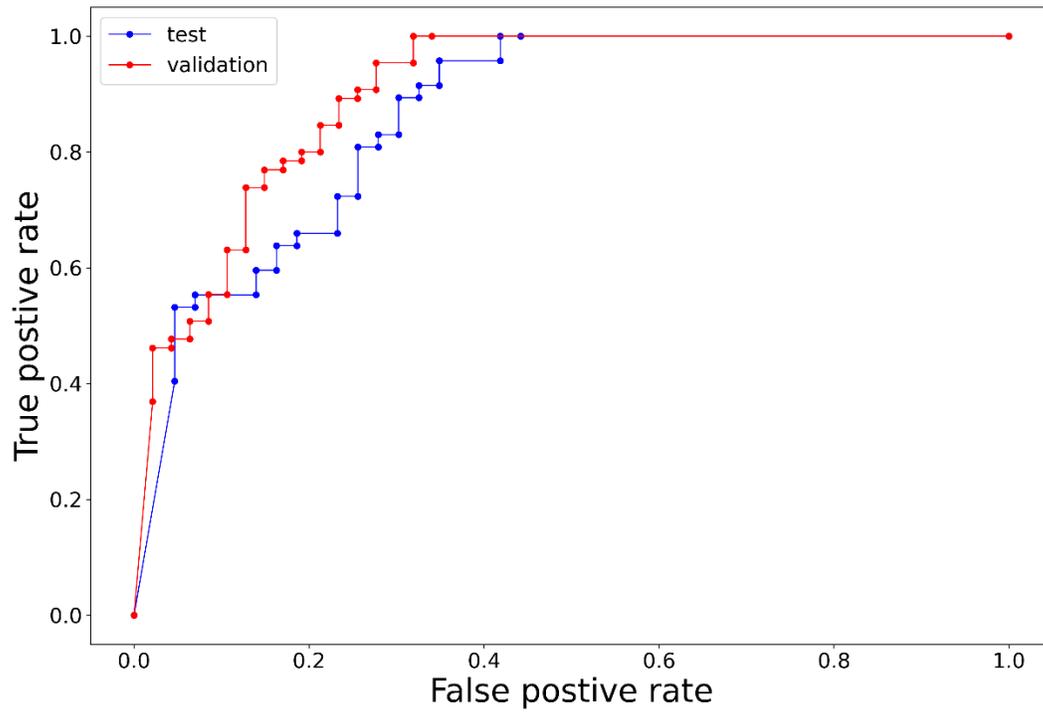


Figure C1. Receiver Operating Characteristic (ROC) curve of k-Nearest Neighbourhood (kNN)

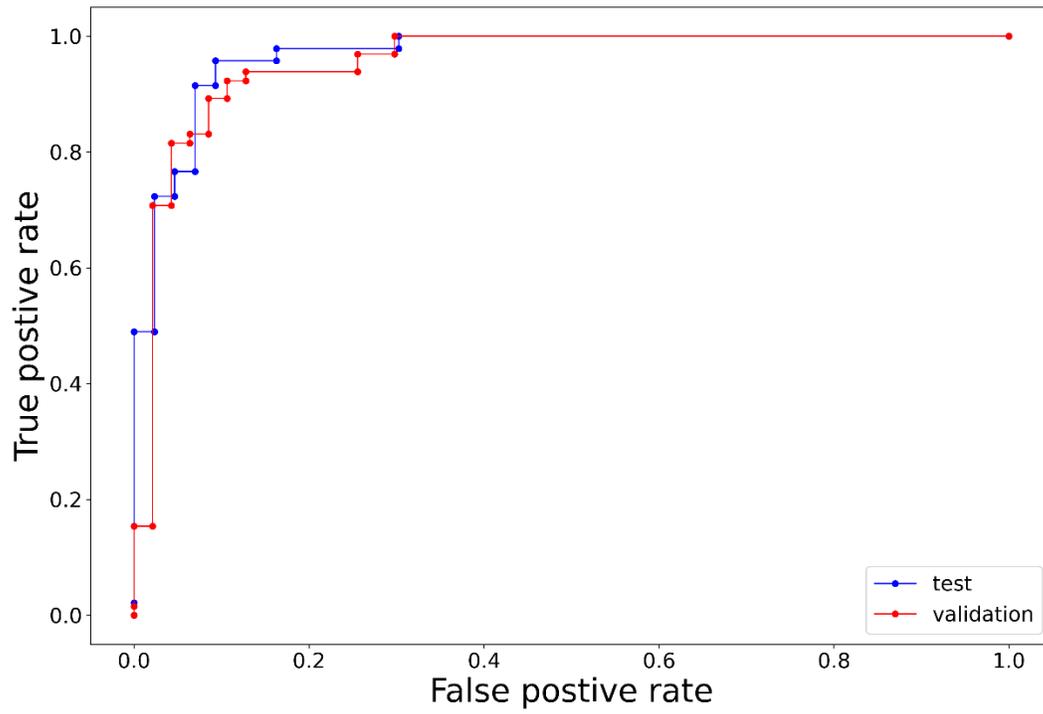


Figure C2. Receiver Operating Characteristic (ROC) curve of Support Vector Classifier (SVC)

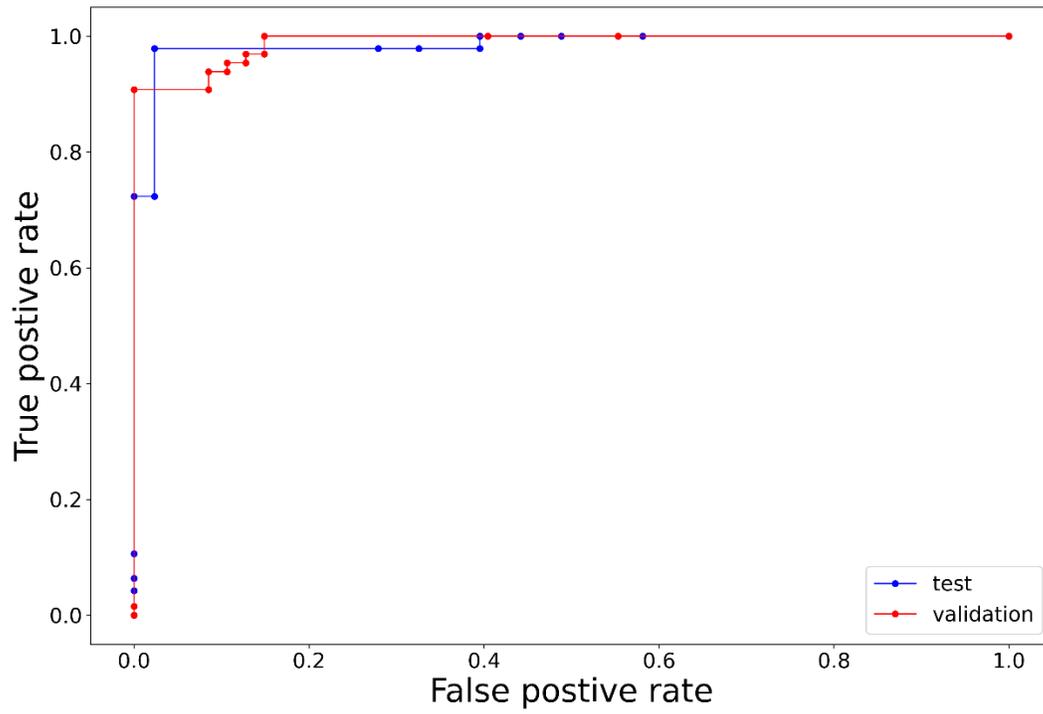


Figure C3. Receiver Operating Characteristic (ROC) curve of Random Forests (RF)

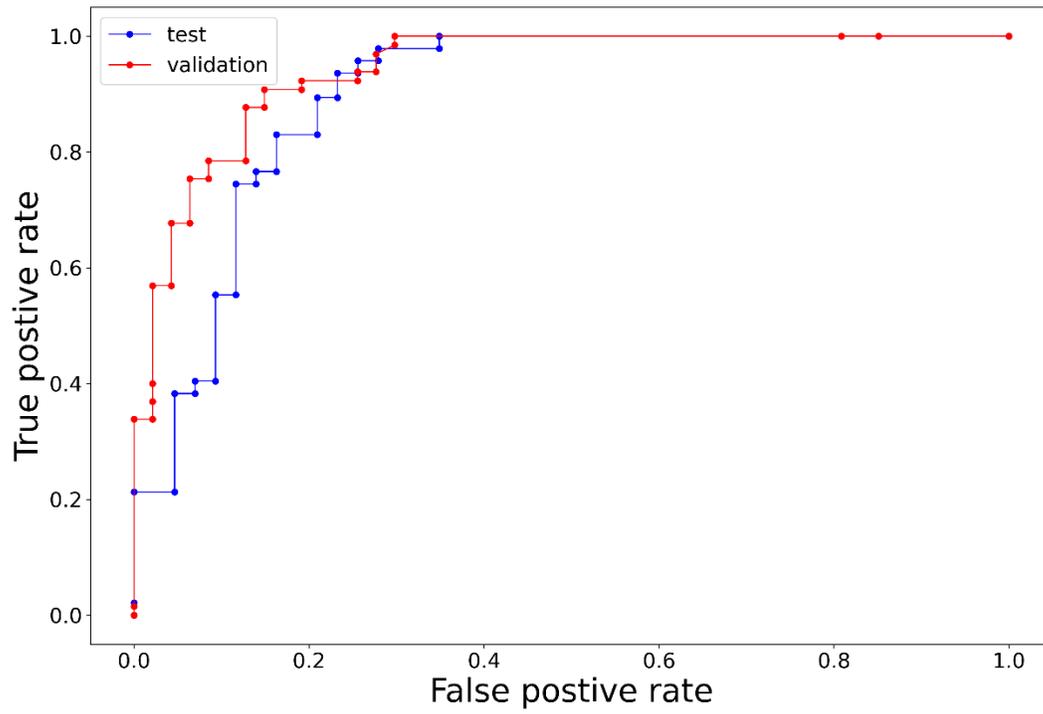


Figure C4. Receiver Operating Characteristic (ROC) curve of Bernoulli Naïve Bayes classifier (NB)

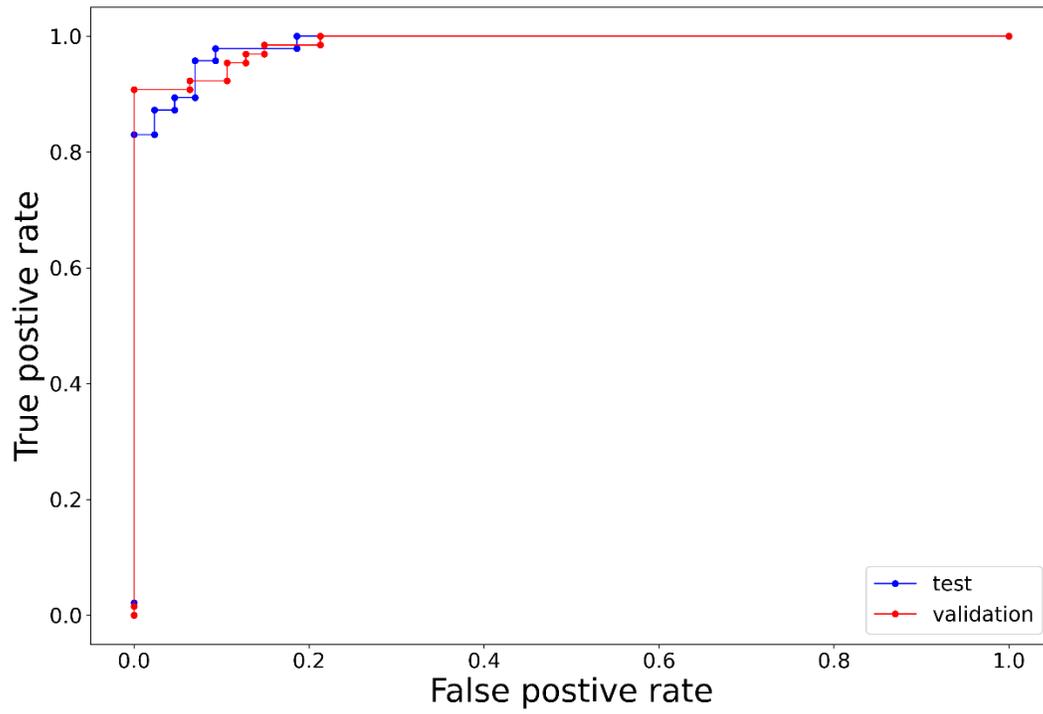


Figure C5. Receiver Operating Characteristic (ROC) curve of Gradient Boosting (GB)

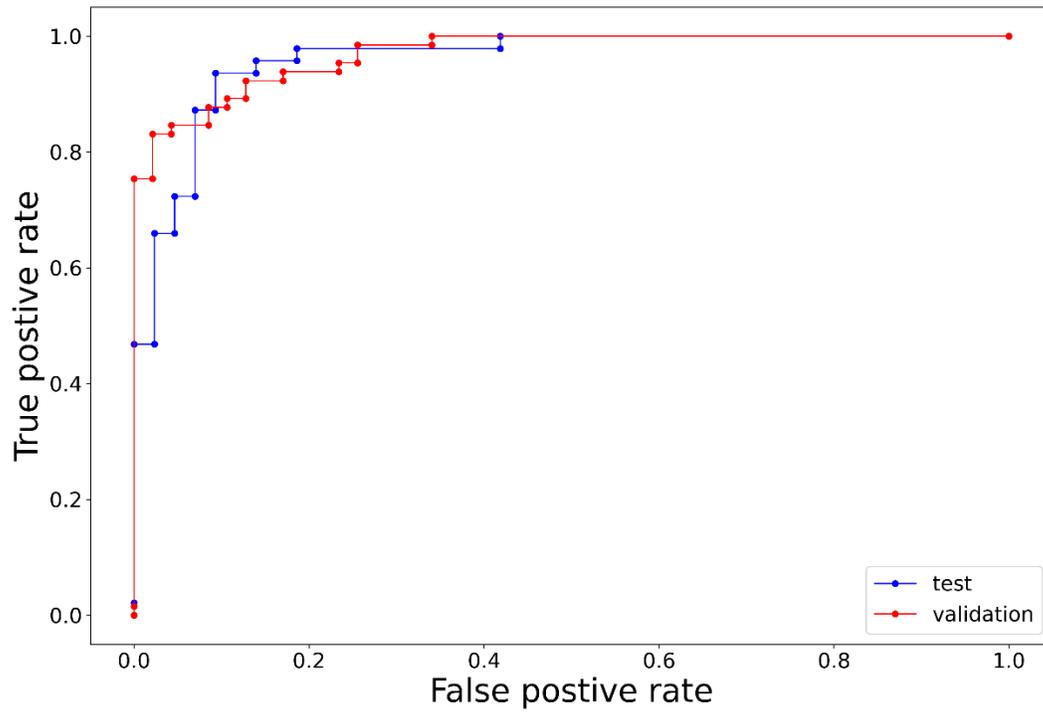
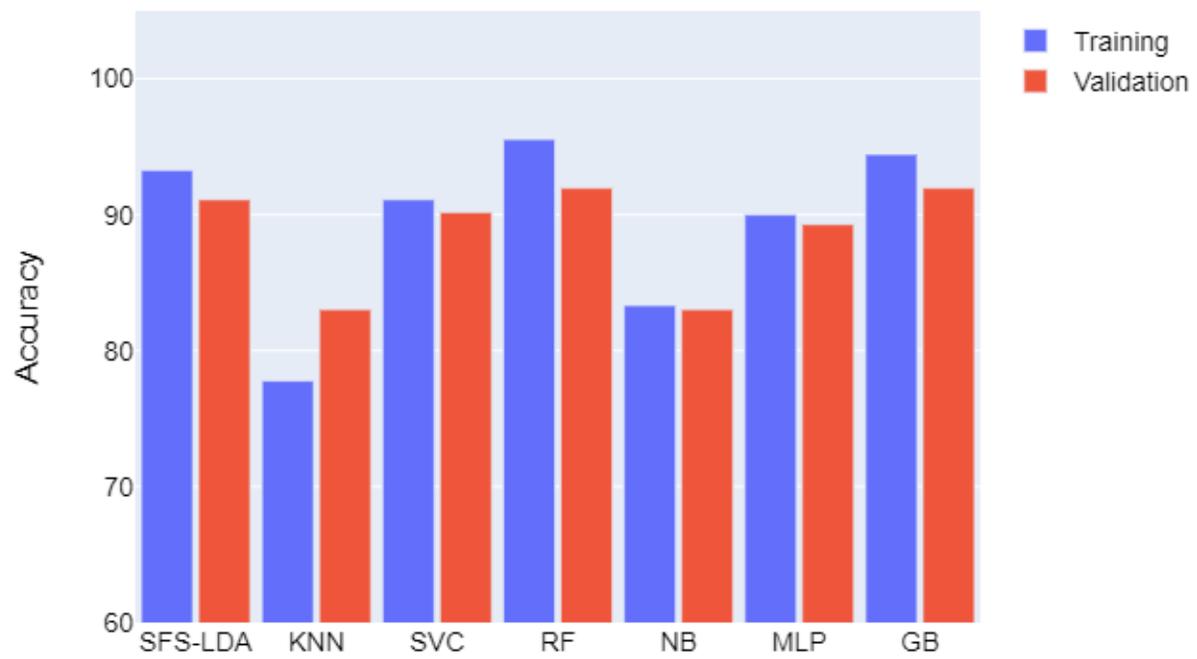
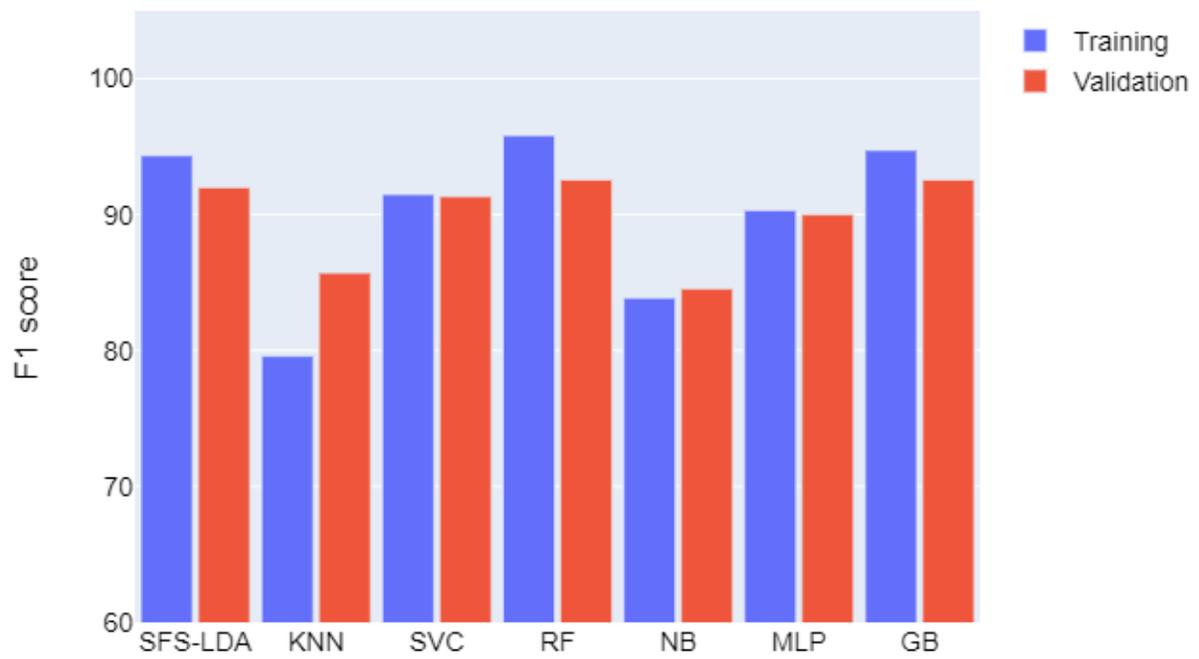


Figure C6. Receiver Operating Characteristic (ROC) curve of Multilayer Perceptron (MLP) neural networks

a.



b.



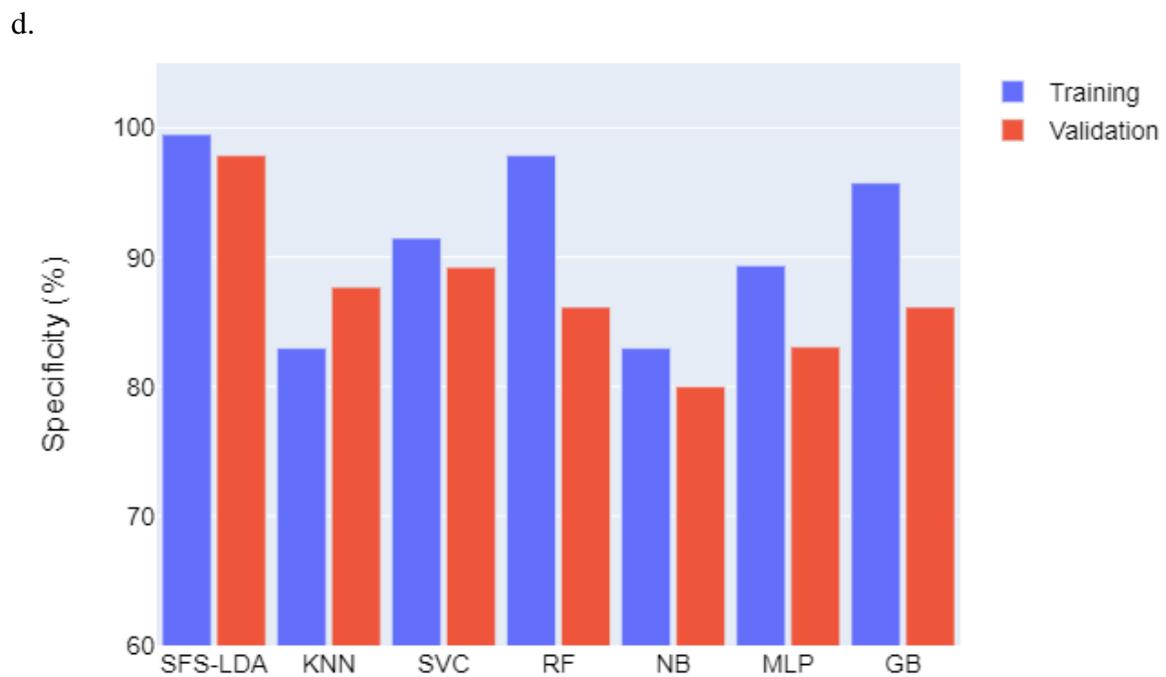
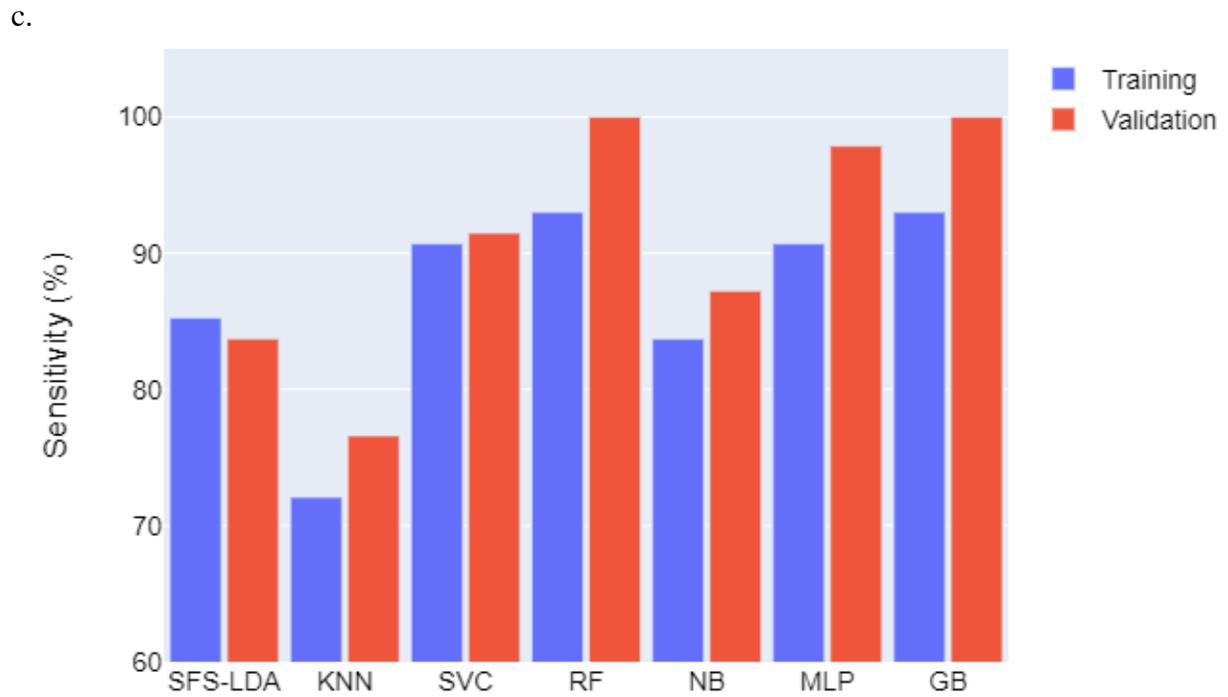


Figure C7. Predicting performance of training and validation datasets in a. Accuracy, b. F1 score, c. Sensitivity, and d. Specificity of different testing models

Table C1 Summary of the hyperparameters for each machine learning model

Models	Hyperparameters
Cross-validation (CV) for grid search and determining model predictability by 10-fold.	
RF	Bootstrap: True
	Criterion: Gini (for Gini impurity), Entropy (for information gain)
	Maximum depth: 10, 30, 50, 70, 90, 100, 200
	Maximum features: Auto, Sqrt, Log2
	Minimum samples leaf: 1, 2, 4
	Minimum samples split: 2, 5, 10
	Number of estimators: 50, 100, 200,500
kNN	Number of neighbours: 1-50
	Weight options: Uniform (for uniform weights), Distance (i.e., weight points by the inverse of their distance)
	Algorithms (used to compute the nearest neighbours): Auto, Ball tree (i.e., BallTree), KD_tree (i.e., KDTree), Brute (i.e., brute-force serch)
Bernoulli NB	Alpha (i.e., smoothing parameter): 1, 0.5, 0.1
	Fit_prior: True
SVC	C(i.e., regularization parameter): 0.1, 1, 10, 100, 1000
	Gamma: 1, 0.1, 0.01, 0.001
	Kernel: RBF, Linear, Poly, Sigmoid
MLP	Hidden layer sizes: 100
	Activation: Identity, Logistic, Tanh, Relu
	Solver: SGD, Adam
	Alpha: 0.0001, 0.001, 0.01, 1
	Learning rate: Constant, Adaptive, Invscaling
GB	Loss: deviance, exponential
	Learning rate: 0.01, 0.05, 0.1, 0.2
	Min samples split: 0.1,0.2,0.3,0.4,0.5
	Minimum samples leaf: 0.1,0.2,0.3,0.4,0.5
	Maximum depth: 3,5,8
	Maximum features: Log2, Sqrt
	Criterion: Friedman MSE, MAE
	Subsample: 0.5, 0.6, 0.8
	Number of estimators: 50,100,200,300

Table C2. BioTox prediction for manufactured and used OECD PFAS in PubChem within the application domain and over 95% probability for classification.

Name	smiles	FA1 score	FA2 score	FA3/4 score	FA5 score	FA6 score	FA7 score	IFABP score	LFABP score	PPAR $\alpha$ score	PPAR $\beta$ score	PPAR $\gamma$ score	Predicted Class
9554	<chem>C(=O)(C(C(C(C(C(C(F)(F)F)(F)(F)F)(F)F)(F)F)(F)F)O</chem>	-8.2	-8.2	-8.2	-9	-8.1	-8.2	-8.7	-9.2	-8.2	-7.6	-8.2	1
67821	<chem>C(=O)(C(C(C(C(C(C(C(F)(F)F)(F)(F)F)(F)(F)F)(F)(F)F)O</chem>	-8.5	-9.1	-8.5	-9.8	-8.5	-8.5	-9.2	-9.7	-8.7	-8	-9.4	1
91708	<chem>C1=CC(=C(C=C1)F)C(=O)NC(=O)NC2=CC(=C(C=C2)Cl)OC3=C(C=C(C=N3)C(F)(F)F)Cl)Cl)F</chem>	-10.9	-10.7	-9.6	-7.6	-9.9	-9.5	-10.9	-10.8	-9.1	-10	-9.5	1
3014940	<chem>C(=O)(C(C(C(C(C(C(F)(F)Cl)(F)F)(F)F)(F)F)(F)F)O</chem>	-8	-8.4	-8.1	-9.2	-8.2	-8	-8.5	-9.2	-7.9	-7.6	-8.3	1
18134	<chem>CCN(CC(=O)O)S(=O)(=O)C(C(C(C(C(C(C(F)(F)F)(F)F)O</chem>	-8.3	-9	-8.3	-9.7	-8.7	-8.3	-9.8	-9.2	-7	-8.4	-7.4	1

	)F)F)F)F) F)F)F)F)												
69785	C(C(C(C(C (F)(F)S(=O ) (=O)N)(F F)(F)F)(F)F ) (C(C(C(F)( F)F)(F)F)(F )F)(F)F	-8.8	-9.4	-8.7	-10.1	-8.4	-8.7	-9.5	-9.3	-8.8	-8.1	-8.5	1
3373415	C(=O)(C(C( C(C(C(C(C (C(F)(F)F)( F)F)(F)F)(F )F)(F)F)(F F)(F)F)(F)F )N	-8.8	-9.3	-8.7	-10	-8.6	-8.7	-9.3	-9.7	-8.9	-8.2	-9.2	1
19232	C1=CC(=C (C2=C1NC (=N2)C(F)( F)F)Cl)Cl	-8.2	-7.5	-6.9	-6.1	-6.9	-6.9	-6.9	-7.3	-6.6	-6.4	-6.4	1
62276	CCOC(=O) C(C)OC(= O)C1=C(C =CC(=C1) OC2=C(C= C(C=C2)C( F)(F)F)Cl)[ N+](=O)[O- ]	-9.5	-9.2	-8.2	-8	-8	-8.2	-9.5	-9.3	-8.3	-8	-8.1	1
87407	CCN1C(=N C2=C1C=C (C=C2)C( F)(F)F)[N+] (=O)[O-]C	-8.3	-7.5	-6.6	-5.7	-6.6	-6.6	-7.2	-7.8	-7.6	-6.4	-6.6	1

2782473	C(=O)(C(C(C(C(C(C(F)(F)Cl)(F)F)(F)F)(F)F)(F)F)(F)F)O	-8.4	-9.2	-8.8	-9.9	-8.3	-8.8	-9	-9.8	-8.5	-8.2	-8.3	1
67787	C1=CC(=C(C(=C1)SC(F)(F)F)[N+](=O)[O-])	-6.6	-6.7	-6.3	-6.7	-6	-6.3	-7.1	-6.6	-5.9	-6.4	-6.6	1
86429	CC(C)N(C1=CC=C(C=C1)F)C(=O)COC2=NN=C(S2)C(F)(F)F	-8.9	-9.3	-7.7	-7.9	-7.5	-7.9	-8.9	-8.5	-7.7	-8.3	-7.2	1
20056430	C1=CC(=C(C(=C1)C(F)(F)F)C(=N)NC(=O)NC2=CC=C(C=C2)OC(F)(F)F)CC3=CC=C(C=C3)C#N	-10.9	-10.9	-9.5	-9.5	-8	-9.4	-11.1	-10.9	-9.1	-10.7	-9.9	1
22286931	CN(CC(=O)O)S(=O)(=O)C(C(C(C(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)F)(F)F)F	-9.1	-9.9	-8.4	-10	-8.9	-8.5	-10.5	-9.1	-7.3	-9.5	-7.4	1
67717	C(=O)(C(F)(F)F)N	-4.1	-4.6	-4.2	-4.3	-4.2	-4.5	-4.6	-4.4	-4.5	-4.5	-4.4	0

67920	C(C(C(C(C(C(F)(F)I)(F)F)(F)F)(F)F)(F)F)(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)	-9	-10	-9.2	-11.2	-9	-9.1	-10.2	-10.2	-8.4	-9.1	-8.8	0
272696	C=C(C(F)(F)F)Br	-4.8	-4.6	-4.3	-4.7	-4.2	-4.3	-5	-4.8	-4.3	-4.3	-4.3	0
10000	C(CCl)C(F)(F)F	-4.5	-4.4	-4.2	-4.4	-4.2	-4.2	-5	-4.6	-4.1	-4.2	-4.4	0
66265	CCNC(C)C C1=CC(=C C=C1)C(F)(F)F	-7.7	-7.9	-6.9	-7.9	-7.1	-6.9	-7.6	-7.6	-6.5	-6.1	-6.8	0
67716	C(=O)(C(F)(F)F)F	-4.3	-4.2	-4.1	-4.4	-4.1	-4.1	-4.9	-4.4	-4.7	-4.4	-4.5	0
120203	C(C(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)	-8.7	-9.9	-9.3	-11	-9	-9.3	-10	-10.2	-8.8	-9	-9.8	0
2736716	C=C(C(F)(F)F)Cl	-4.7	-4.6	-4.2	-4.6	-4.3	-4.2	-5	-4.8	-4.2	-4.4	-4.3	0
3226	C(C(OC(F)F)(F)F)(F)C I	-5.4	-5.1	-5.2	-5.5	-5.2	-5.2	-5.7	-5.3	-5.1	-5.2	-5.2	0
17822	C(C(F)(F)F)(F)Cl	-4.5	-4.5	-4.2	-4.7	-4.3	-4.2	-5	-4.9	-4.2	-4.3	-4.4	0
3763	C(C(F)(F)F)(OC(F)F)C I	-5.3	-5.2	-4.9	-5.5	-5.1	-4.9	-5.7	-5.3	-5.2	-5.2	-4.9	0
9385	C(C(F)(F)F)(Cl)Cl	-4.5	-4.5	-4.1	-4.6	-4.3	-4.1	-5	-4.7	-4.2	-4.3	-4.5	0
9633	C(C(F)(F)F)(F)F	-4.5	-4.4	-4.1	-4.6	-4.2	-4.3	-5	-4.8	-4.2	-4.2	-4.4	0
9893	COC(=O)C(F)(F)F	-4.4	-4.3	-4.1	-4.2	-4.2	-4.3	-4.7	-4.5	-4.6	-4.6	-4.6	0

61106	C(=O)(C(F)(F)F)Cl	-4.4	-4.3	-4.1	-4.3	-4.2	-4.1	-4.8	-4.5	-4.5	-4.5	-4.4	0
5709018	C(=CCl)C(F)(F)F	-4.6	-4.6	-4.3	-4.6	-4.2	-4.3	-5.3	-4.8	-4.1	-4.3	-4.5	0
6422	C(=O)(C(F)(F)F)O	-4.1	-4.4	-4.1	-4.3	-4.1	-4.4	-4.5	-4.3	-4.4	-4.6	-4.5	0
9844	C=COCC(F)(F)F	-4.5	-4.2	-4.3	-4.4	-4.1	-4.3	-5	-4.6	-4.1	-4.4	-4.5	0
67900	C(F)(F)(F)S Cl	-4	-3.8	-3.7	-4.1	-3.8	-3.8	-4.3	-4.1	-3.7	-3.9	-3.9	0
67901	C(C(F)(F)F)(O)O	-4	-4.7	-4	-4.4	-3.9	-4.5	-4.3	-4.3	-4.2	-4.7	-4.5	0
3337	CCNC(C)C C1=CC(=C C=C1)C(F) (F)F	-7.7	-7.9	-6.9	-7.9	-7.1	-6.9	-7.6	-7.6	-6.7	-6.9	-6.9	0
3562	C(C(F)(F)F)(Cl)Br	-4.6	-4.4	-4.2	-4.7	-4.3	-4.2	-5	-4.7	-4.2	-4.3	-4.2	0
4116	COC(C(Cl)Cl)(F)F	-4.5	-4.4	-4.2	-4.8	-4.3	-4.4	-4.9	-4.7	-4.4	-4.5	-4.7	0
9639	C(C(C(C(F)(F)F)(F)F)(C(C(F)(F)F)(F)F)(F)F)	-8.3	-7.9	-7.4	-8.5	-7.6	-7.4	-8.1	-7.8	-7.7	-8.1	-7.9	0
9871	CC(=O)C(F)(F)F	-4.5	-4.3	-4.1	-4.3	-4.2	-4.2	-4.9	-4.6	-4.4	-4.5	-4.4	0
9892	COC(C(F)(F)F)O	-4.2	-4.4	-4.2	-4.5	-4.1	-4.4	-4.7	-4.3	-4.5	-4.5	-4.5	0
12672	C=CC(F)(F)F	-4.3	-4.2	-3.9	-4.3	-4.1	-3.9	-4.8	-4.6	-3.9	-4	-4.2	0
137512	C(CO)C(F)(F)F	-4.3	-4.6	-4.2	-4.2	-4.2	-4.4	-4.7	-4.6	-4.1	-4.2	-4.4	0
11116025	C(=CF)C(F)(F)F	-4.7	-4.4	-4.2	-4.5	-4.3	-4.2	-5.2	-4.8	-4.1	-4.4	-4.6	0
6393	C(F)(F)(F)F	-3.6	-3.5	-3.4	-3.6	-3.3	-3.4	-4	-4.1	-3.2	-3.5	-3.6	0
9868	CC(F)(F)F	-3.7	-3.6	-3.5	-3.7	-3.4	-3.5	-4.2	-4	-3.4	-3.6	-3.8	0
6409	C(C(F)(F)F)O	-3.7	-3.9	-3.7	-4.1	-3.8	-3.9	-4	-3.9	-3.9	-3.9	-4	0

9773	C(C(F)(F)F) N	-3.6	-4	-3.7	-4.1	-3.8	-3.9	-3.9	-4	-3.6	-3.8	-3.8	0
13129	C(C(F)(F)F) F	-3.9	-3.8	-3.6	-4	-3.8	-3.7	-4.4	-4.2	-3.7	-3.7	-3.9	0
67710	C(#N)C(F)( F)F	-4	-3.8	-3.8	-3.8	-3.8	-3.8	-4.2	-4.1	-3.6	-3.7	-4.1	0
9774	CC(C(F)(F) F)O	-4.3	-4.4	-4.4	-4.3	-4.2	-4.2	-4.7	-4.5	-4.4	-4.5	-4.3	0
65564	C(=O)C(F)( F)F	-3.7	-3.6	-3.5	-3.9	-3.6	-3.6	-4.1	-3.9	-4	-4	-4	0
6408	C(C(F)(F)F) Cl	-4	-3.8	-3.7	-4.1	-3.9	-3.8	-4.4	-4.2	-3.6	-3.8	-4	0
62406	C(F)(F)(F)S (=O)(=O)O	-4.3	-4.8	-4.4	-4.5	-4.3	-4.4	-4.8	-4.7	-4.7	-5	-4.7	0
67899	CCC(F)(F) F	-4.3	-4.2	-3.9	-4.3	-4.1	-3.9	-4.8	-4.7	-3.9	-4	-4.2	0
5708720	C(=CF)C(F) (F)F	-4.7	-4.4	-4.3	-4.5	-4.3	-4.2	-5.2	-4.8	-4.1	-4.4	-4.6	0