About author manuscripts | Submit a manuscript

# Integrative computational approaches to inform relative bioaccumulation potential of per- and polyfluoroalkyl substances (PFAS) across species

**Weixiao Cheng**[1], **Jon A. Doering**[2], **Carlie LaLone**[3,*], **Carla Ng**[1,4,*]

[1]Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA

[2]National Research Council, Duluth, MN

[3]Great Lakes Toxicology and Ecology Division, Center for Computational Toxicology and Exposure, Office of Research and Development, US Environmental Protection Agency, Duluth, MN

[4]Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA

## Abstract

Predictive toxicology is increasingly reliant on innovative computational methods to address pressing questions in chemicals assessment. Of importance is the evaluation of contaminant impact differences across species to inform ecosystem protection and identify appropriate model species for human toxicity studies. Here we evaluated two complementary tools to predict cross-species differences in binding affinity between per- and polyfluoroalkyl substances (PFAS) and the liver fatty acid binding protein (LFABP): the Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) tool and molecular dynamics (MD). SeqAPASS determined that the structure of human LFABP, a key determinant of PFAS bioaccumulation, was conserved in the majority of vertebrate species, indicating these species would have similar PFAS bioaccumulation potentials. Level 3 SeqAPASS evaluation identified several potentially destabilizing amino acid differences across species, which were generally supported by DUET stability change predictions. Nine single-residue mutations and seven whole species sequences were selected for MD evaluation. One mutation (F50V for PFNA) showed a statistically significant difference with stronger affinity than wild-type human LFABP. Predicted binding affinities for 9 different PFAS across 7 species showed human, rat, chicken and rainbow trout had similar binding affinities to one another for each PFAS, whereas Japanese medaka and fathead minnow had significantly weaker LFABP binding affinity for some PFAS. Based on these analyses, the combined use of SeqAPASS and molecular dynamics provides rapid screening for potential species differences with deeper structural insight. This approach can be easily extended to other important biological receptors and potential ligands.

*Co-corresponding authors Contact author: Carla Ng, University of Pittsburgh, Civil & Environmental Engineering, 3700 O'Hara St, Pittsburgh, PA, 15261. carla.ng@pitt.edu.

SUPPLEMENTARY DATA

Supplementary data are available at Toxicological Sciences online.

**Keywords**

PFAS; species differences; molecular modeling; SeqAPASS

---

# INTRODUCTION

Per- and polyfluoroalkyl substances (PFAS) encompass a class of potentially thousands of short-chained, long-chained, and branched organofluorine structures (e.g., perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA)) (Wang et al. 2017). These synthetic chemicals have been used in numerous industrial applications and consumer products including fire-fighting foams, as stain and oil repellents, in lubricants, apparels, upholstery, etc. (ITRC 2020). Their wide-spread use and persistence in the environment, as well as their ability to bioaccumulate have been recognized globally, making PFAS of research interest, particularly in understanding potential toxicities across species. Due to the ubiquitous nature of PFAS, they have been measured in tissues from species as diverse as whales, birds, fish, and even invertebrates, covering the range of trophic levels (Burkhard 2020). Reproductive toxicity, neurotoxicity, hepatotoxicity, immunotoxicity and modulation of metabolism have been reported as adverse effects of PFAS exposure in mammals and possibly other organisms (Sunderland et al., 2019). However, considering the variety of PFAS, limited studies are available to understand adverse effects, with even less known regarding effects across taxa.

Notably, some PFAS have high bioaccumulation potential in animals. Studies have found that many long-chain PFAS (perfluoroalkyl carboxylic acids (PFCA) with >= 7 perfluorinated carbons and perfluoroalkane sulfonic acids (PFSA) with >= 6 perfluorinated carbons) accumulate in blood, liver, and kidney (Ng and Hungerbühler 2014); their biological half-lives were estimated to be several years for humans (Olsen et al. 2007) and several days for male rats (Kennedy et al. 2004; Kim et al. 2016; Kudo and Kawashima 2003). The underlying molecular mechanisms for PFAS bioaccumulation potential are closely related to protein-PFAS interactions. PFAS have high binding affinity to serum albumin and to liver-type fatty acid binding proteins (LFABP) in liver and kidney tissues, making those tissues important accumulation media (Han et al. 2003; Sheng et al. 2018; Woodcroft et al. 2010). Moreover, cellular transport of PFAS is most likely controlled by both passive diffusion and active transport facilitated by transporter proteins (Weaver et al. 2009; Yang et al. 2009; 2010), which impacts species-specific elimination half-lives. By explicitly considering key proteins and transporters, we developed a physiologically based toxicokinetic model that successfully simulated the toxicokinetics and tissue distribution of PFAS in rat and fish (Cheng and Ng 2017; Ng and Hungerbühler 2013). Both experimental studies and modeling results demonstrate that protein-PFAS interactions play essential roles in determining PFAS bioaccumulation potential in animals. Thus, estimates of the strengths of these interactions can be used as proxies to evaluate their bioaccumulation potential.

Due to the large number of PFAS involved and potential species of interest, with limited resources for testing, it is not feasible to assess the bioaccumulation potential for all PFAS and all species through laboratory experiments. *In silico* approaches, on the other hand, hold

great promise for hazard and risk assessment. Here, we propose an integrative *in silico* approach to inform PFAS bioaccumulation potential across different species from the perspective of the sequence, structure, and function of a critical protein, LFABP. There are two main components included in our approach. The first is the US Environmental Protection Agency's Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS; seqapass.epa.gov/seqapass/; v4.0), a web-based tool used for cross-species extrapolation for chemical toxicity based on assessment of protein sequence and structural similarity (LaLone et al. 2016). It consists of three levels of sequence evaluation, ranging from the whole primary amino acid sequence alignment to conserved functional domain and individual residue alignments, with increasing levels of complexity (LaLone et al. 2016). The second is a multi-step molecular modeling workflow that combines homology modeling, molecular docking, and molecular dynamics (MD) to estimate the protein binding affinities for different PFAS, which can provide additional structural insights. Homology modeling can predict protein structures based on sequences and is used to construct the 3-dimensional structure for proteins that have no structural information available (Kelley et al. 2015). Molecular docking, a powerful tool to predict the conformation of ligands bound to proteins (Trott and Olson 2010), is then used to generate protein-PFAS complex structures. Taking the initial complex structure as input, MD simulation was used to estimate the protein binding affinity for each PFAS. Our previous study has shown that the MD approach can predict the relative protein binding affinity of different PFAS structures in a fast and reliable way (Cheng and Ng 2018).

To demonstrate the utility of this integrative *in silico* approach, we focused on LFABP as the protein proxy for bioaccumulation assessment in this study. LFABP is a well-studied protein thought to explain the high accumulation of PFAS in liver tissue, and there are many available experimental binding affinity data for different PFAS, which we previously used for evaluation of the multi-step molecular modeling workflow (Cheng and Ng 2018). Here, we considered a total of 9 PFAS with different functional head groups and fluorinated carbon chain lengths including 6 PFCAs (i.e., PFBA, PFPA, PFHxA, PFHpA, PFOA, and PFNA) and 3 PFSAs (i.e., PFBS, PFHxS, and PFOS), the chemical structure and other information for those PFAS are summarized in Table 1 and in Supplementary Data, Table S1.

## MATERIALS AND METHODS

### SeqAPASS Workflow to Predict PFAS Susceptibility across Species.

The US Environmental Protection Agency Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS; https://seqapass.epa.gov/seqapass/; v4.0 data version 4) tool was used to evaluate protein conservation and predict bioaccumulation potential across species for different PFAS (LaLone et al. 2016). The SeqAPASS evaluation was also used to predict which critical amino acid residues could be influencing potential differences in bioaccumulation across species by comparing amino acid side chain classification and molecular weight among species-specific amino acid substitutions (Doering et al. 2018) The query protein used for evaluating primary amino acid sequence conservation using SeqAPASS Level 1 was the human LFABP (NCBI Reference Sequence accession

NP_001434). No functional domains were identified as specific hits in NCBI's Conserved Domains database (Marchler-Bauer et al. 2015), therefore, no SeqAPASS Level 2 runs were submitted. Previous computational studies employing molecular docking and molecular dynamics simulations were conducted to evaluate the interactions between human and Norway rat LFABP with PFAS. A number of amino acids were identified as important because they formed hydrogen bonds with the PFAS and/or had large energy contributions. These hydrogen-bond-forming amino acids were therefore identified as critical amino acids for evaluation in SeqAPASS Level 3 and were used to predict conservation of the LFABP sequence across species. These previous studies also indicate that there are differences in binding affinity between human and rat (Cheng and Ng 2018). To further explore these species' differences and to identify other potentially important amino acids, the human LFABP was aligned with the Norway rat LFABP (NCBI Reference Sequence accession NP_036688.1) using NCBI Constraint-based Multiple Alignment Tool, where 22 of the 127 amino acids were not exact matches (Supplemental Data, Figure S1). All 22 individual amino acids were evaluated in SeqAPASS Level 3, with only 6 amino acids predicted to impact binding considering human and rat LFABP, specifically positions 48, 50, 54, 81, 97, and 104 (using human LFABP as the template). Conservation of these six amino acid positions was then evaluated further across vertebrate LFABP sequences using SeqAPASS Level 3, which compares amino acid side chain classification and molecular weight (differences of >30g/mol) to predict cross species conservation in protein-chemical interaction (Doering et al. 2018). Amino acid differences in these positions were identified across taxa and submitted to DUET (http://biosig.unimelb.edu.au/duet/stability), which is a web-tool for predicting the effects of mutations on protein stability (Pires et al. 2014). DUET was used to predict whether the changes in amino acids (representing those seen in other species) altered stability by submitting the crystal structure for human LFABP as input with specified mutations.

## Multi-step Molecular Modeling to Predict Protein Binding Affinity.

**Multi-step Molecular Modeling Workflow.—**The multi-step molecular modeling workflow was developed based on our previous molecular dynamics modeling to estimate the LFABP binding affinity for different PFAS. As shown in Figure 1, the workflow consists of four major steps: curation of structures, molecular docking, molecular dynamics, and molecular mechanics combined with Poisson-Boltzmann surface area (MM-PBSA) calculation.

Molecular structures were either the 3-dimensional crystal structures obtained from the Protein Data Bank (PDB, https://www.rcsb.org/) or were constructed by the homology modeling tool Phyre2 for LFABP across different species. In the PDB, we chose protein structures based on their resolution (the higher, the better) and completeness of key residues. Phyre2 was used to construct the 3-dimensional structure because it is one of the most popular protein structure prediction servers and very user-friendly (Kelley et al. 2015). We selected the structure with the highest confidence as the output of Phyre2. All protein structures constructed for this study have a confidence of 100%. The 3-dimensional structures for PFOA and PFOS were obtained from the PDB (PDB codes 5JID and 4E99, respectively), and for other PFAS ligands were constructed using Avogadro (v1.2.0)

(Hanwell et al. 2012). For each ligand generated in Avogadro, molecular mechanics-based geometry optimization was performed to ensure a realistic rendition of the molecule (Hanwell et al. 2012). To evaluate the performance of Avogadro in generating PFAS structures when experimental ones are not available, the 3-dimensional structures of PFOA and PFOS were also created and compared with those obtained from the PDB. As shown in Supplemental Data Figure S2, the root-mean-square deviation (RMSD) between those two versions of structures is 0.897 Å for PFOA and 0.692 Å for PFOS. Those values demonstrate the effectiveness of using Avogadro to generate 3-dimensional structures for PFAS (Morris and Lim-Wilby 2008).

Next, the molecular docking tool Autodock Vina (v1.1.2) (Trott and Olson 2010) was employed to generate the initial structures of protein-PFAS complexes following the same procedure we developed in our previous study (Cheng and Ng 2018). The outputs from the docking experiments include binding free energies and docking poses for each protein-PFAS pair. The top 3 strongest binding modes (i.e., the lowest energies) were selected as initial structures for molecular dynamics (MD) simulations.

The MD simulations of all protein-PFAS complexes were performed with the AMBER 14 suite (Case et al. 2014), as described in our previous study (Cheng and Ng 2018). Briefly, the protein-PFAS complex system was first explicitly solvated in a cubic box of TIP3P water molecules (Jorgensen et al. 1983), with the addition of Na+ or Cl− counterions to neutralize the systems. Next, the whole system was subject to a series of processes to mimic experimental conditions (e.g., at constant pressure). Those processes include: (1) 3500 cycles minimization; (2) heating with constant volume and temperature for 20 ps from 0 K to 300 K; (3) adjusting density to 1 $g/cm^3$ at constant pressure (1 bar) for 100 ps; (4) equilibration with constant pressure (1 bar) and temperature (300 K) for 2 ns; (5) production with constant pressure (1 bar) and temperature (300 K) for 24 ns. Each phase generates trajectories containing coordinates and velocity information of the molecular system, which can be used to calculate free energy of binding ( $G_{bind}$).

Finally, the MM-PBSA method (Miller III et al. 2012) was used to calculate $G_{bind}$ as follows:

$$\Delta G_{bind} = G^{Complex} - G^{Protein} - G^{PFAS}$$

where $G^{Complex}$, $G^{Protein}$, and $G^{PFAS}$ are the free energies of complex, proteins, and PFAS ligands, respectively. The free energy (G) of each state was estimated from the following sum (Homeyer and Gohlke 2012):

$$G = <E_{bond} + E_{el} + E_{vdw} + G_{polar} + G_{nonp} - TS>$$

where the brackets indicate an average over MD trajectories. Inside the brackets, the first three terms are molecular mechanical energy terms for bonded, electrostatic and van der Waals interactions, respectively. $G_{polar}$ and $G_{nonp}$ are the polar solvation free energy and the nonpolar contribution, respectively. The last term is the absolute temperature (T) multiplied

by the entropy (S). All those terms can be calculated based on the trajectories obtained from MD simulation.

The relationship between $\Delta G_{bind}$ (unit in kcal/mol) and binding affinity which is quantified by the equilibrium dissociation constant ($K_d$, unit in μM) is as follows (Caldwell and Yan 2004; Kastritis and Bonvin 2013):

$$\Delta G_{bind} = RT ln \frac{K_d}{c_0}$$

where R is gas constant (1.987 cal $K^{-1}$ $mol^{-1}$), T is temperature (300 K), and $c_0$ is the standard state concentration (1 M). The equation above shows that a lower $\Delta G_{bind}$ value means a stronger binding affinity.

**Single Residue Mutation Effects.**—SeqAPASS Level 3 results identified a number of key amino acid residues differing between human and other species, and those residues include F50, A54, T81, T93, and N97. To further evaluate how the mutation of those residues would affect protein binding affinity to PFAS, the multi-step molecular modeling was used to estimate the binding affinity of mutant human LFABP for PFAS. Based on SeqAPASS Level 3 results, we considered a total of 8 single amino acid substitutions including F50V, F50I, F50L, A54T, T81A, T81G, T93A and N97G (here, F50V means residue at position 50 mutated from F to V, the same for other mutations), and 1 double amino acid substitution 5093 which indicates two sequential mutations F50V, followed by T93A. The 3-dimensional structures of those mutant human LFABPs were generated using DUET and then fed into the molecular modeling workflow. The calculated binding affinities for those mutations were compared with the wild type human LFABP to assess the mutation effects on protein binding affinity for PFAS. Finally, to evaluate whether this methodology works, experimental data from the Sheng et al. study (Sheng et al. 2016) was used to compare with the model results. In the Sheng et al. study, the binding affinities of wild-type human LFABP and its variants (i.e., S39G, M74G, N111D, and R122G) for PFOA and PFNA were determined by fluorescence displacement and isothermal titration calorimetry experiments. To compare, the binding affinities for S39G, M74G, N111D, and R122G mutation were predicted using the molecular modeling workflow.

**Whole Protein Cross-species Effects.**—Based on the SeqAPASS Level 1 results, a total of 7 different species (i.e., human, rat, chicken, zebrafish, rainbow trout, Japanese medaka, and fathead minnow) were selected to further examine the difference of protein binding affinity across species using the full LFABP gene sequence for each species and 9 different PFAS. The 3-dimensional crystal structures were obtained from the PDB website for human LFABP (PDB code: 3STM) and rat LFABP (PDB code: 1LFO). For other LFABPs, Phyre2 (Kelley et al. 2015) was used to construct 3-dimensional structures. The protein sequences used to build the 3-dimensional structures are shown in Supplemental Data, Figure S3. Those molecular structures were then used to predict $\Delta G_{bind}$ values for all LFABP-PFAS pairs from the molecular modeling workflow.

**Data Analysis.—**For each LFABP-PFAS pair, a total of 9 $G_{bind}$ values (those values were calculated based on the trajectories from 9 independent simulation phases and thus can be considered as random samples) were generated from the molecular modeling workflow. One-way ANOVA was conducted to test for significant differences among the 7 different species of LFABP for the 9 tested PFAS ligands. In addition, multiple comparisons with Tukey's test was performed to identify which groups are significantly different from each other for both single residue mutation effects and whole protein cross-species effects. The Python package SciPy (https://scipy.org/index.html) and statsmodels (https://github.com/statsmodels/statsmodels) were used for ANOVA and Tukey's test, respectively; and both tests were conducted based on the 9 different $G_{bind}$ values for each LFABP-PFAS pair.

## RESULTS

### SeqAPASS Level 1: Conservation of Primary Amino Acid Sequence

The SeqAPASS Level 1 primary amino acid sequence comparison of human LFABP resulted in the alignment of sequences from 347 species (62 of which were identified as ortholog candidates), across 21 taxonomic groups representing vertebrates, invertebrates, and fungi. Results from SeqAPASS predict that LFABP is conserved in 302 species from Mammalia, Aves, Lepidosauria, Testudines, Crocodylia, Amphibia, Actinopteri, and Chondrichthyes taxonomic groups, whereas 43 invertebrates and fungi lack conservation of the protein (Figure 2; (Supplemental Data SeqAPASS Output S1). Therefore, SeqAPASS Level 1 provides a line of evidence that vertebrates would share similar bioaccumulation potential to PFAS as humans.

### SeqAPASS Level 3: Conservation of Critical Individual Amino Acid Residues

In evaluating critical amino acid residues identified from previous work as involved in hydrogen bonding interactions with PFAS, the SeqAPASS Level 3 prediction aligned 277 vertebrate species (removing hypothetical, partial, and nonmatching annotated proteins from Level 1) to the human LFABP as the template, identifying only 42 species predicted to differ in bioaccumulation potential from that of humans (Supplemental Data SeqAPASS Output S1). Specifically, those that were predicted to differ included eight species from Mammalia, seventeen from Aves, one from Lepidosauria, one from Amphibia, fourteen from Actinopteri (including common model organisms in ecotoxicology Japanese medaka and fathead minnow), and one from Chondrichthyes (Supplemental Data SeqAPASS Output S1). Therefore, the LFABP for the majority of vertebrate species was conserved and predicted to have similar bioaccumulation potential to PFAS as that of humans.

### SeqAPASS Level 3: Identification of Other Potential Critical Amino Acids Across Species

Since Level 3 of the SeqAPASS tool can be used to both extrapolate from known critical amino acids and develop research hypothesis by identifying potential amino acid differences across species that may help explain differences in susceptibility, the SeqAPASS Level 3 evaluation was expanded via comparison of human and rat sequences and then further across taxa (Supplemental Data, S1). The SeqAPASS Level 3 results identified other potentially critical individual amino acid residues in human LFABP—namely, phenylalanine (F50), alanine (A54), threonine (T81), threonine (T93), and asparagine (N97)—were compared to

other vertebrate species identifying four species groups (Table 2). Most primates, ruminants, and whales/dolphins aligned identical amino acids to the human. Rodents and other mammals, fish, amphibians, and testudines have amino acid substitutions aligning with human at positions 50, 54, 81 and 97. Aves, Lepidosauria, and Chondrichthyes have amino acid substitutions aligning with each critical position identified in humans, whereas Crocodylia has amino acid substitutions aligning with human positions 54, 93, and 97. Interestingly, the zebrafish, a common model organism, aligned an alanine at human position T93, whereas all other fish species aligned either a Threonine (no difference from human) or Valine.

These potentially important amino acid differences observed across species by SeqAPASS were also evaluated in DUET for predicted changes in protein stability. In instances where amino acid substitutions were predicted by SeqAPASS to lead to differences in species ability to interact with PFAS, DUET provided another line of evidence that mutations to the human LFABP based on known cross-species substitutions destabilized the protein (Table 2). The only result from SeqAPASS (sequence-based prediction) that differed from DUET (structure-based prediction) was the N97G substitution which was predicted as different from human in SeqAPASS but stabilizing from DUET. This difference suggests that structural consideration can add unique information to the cross-species comparison of changes in amino acids.

### Effect of Single Residue Mutations on Protein Binding Affinity by MD

To evaluate the effectiveness of the MD workflow for predicting the effects of mutation on protein binding affinity, we compared our model predictions to experimental observations by Sheng et al (Sheng et al. 2016). In that study, fluorescence displacement assays showed that after mutations S39G and M74G, the binding affinities for PFOA and PFNA were comparable to wild-type human LFABP (no substantial change), while after N111D or R122G mutations, the binding of both PFOA and PFNA to human LFABP were not detected (loss of binding). Isothermal titration calorimetry experiments in the same study indicated similar results, except that the R122G mutation did not cause a significant change to the human LFABP binding affinity for either PFOA or PFNA. Our molecular modeling results (Figure 3) showed a significant decrease in binding affinity after the R122G mutation for both PFOA and PFNA, whereas S39G and M74G mutations did not change the binding affinity significantly for either PFOA or PFNA. Finally, the N111D mutation caused a significant decrease in predicted binding affinity for PFOA, but not for PFNA (Figure 3). This shows good agreement between the experimental data and the prediction results and demonstrates the capability of our molecular modeling workflow to predict the mutation effects on protein binding affinity for PFAS.

Figure 4 shows the comparison between wild-type human LFABP and mutated LFABP selected based on SeqAPASS results. As indicated, there is no significant difference (P > 0.05) between wild-type human LFABP and all mutations for both PFOA and PFNA, except the single mutation F50V for PFNA, which has a significantly stronger binding affinity than wild-type human LFABP (P < 0.05).

## Cross-species Effects on Protein Binding Affinity using Full Sequences

The $\Delta G_{bind}$ value and its five energy terms (i.e., van der Waals, electrostatic, polar and nonpolar solvation energy, and entropy) were generated from the multi-step molecular modeling workflow for each LFABP-PFAS complex using full gene sequences for human, rat, chicken, rainbow trout, zebrafish, Japanese medaka and fathead minnow (Tables S2-S8 and Figures S3-S6 in Supplemental Data). The average $\Delta G_{bind}$ values over the 9 tested PFAS ligands for human, rat, chicken, and rainbow trout are smaller than −8.0 kcal/mol. This is a substantially lower value (i.e., stronger binding affinity) than that predicted for Japanese medaka and fathead minnow (average $\Delta G_{bind}$ values larger than or equal to −5.25 kcal/mol, Table 3). The binding affinity for zebrafish is between these two groups, with an average $\Delta G_{bind}$ value of −6.44 kcal/mol. A one-way ANOVA shows there is a significant difference across 7 species for all 9 tested PFAS in terms of their LFABP binding affinity, with P values ranging from 1.02E-10 to 0.021 (Table 4).

The multiple comparison Tukey test between human LFABP and the LFABPs for the other 6 species shows that Japanese medaka has significantly weaker LFABP binding affinity compared to human for all PFAS ligands (P < 0.05) except PFHxA, PFOA and PFNA (Figure 5). Fathead minnow also shows significantly weaker LFABP binding affinity than human for PFBS and PFHxS (P < 0.05), while LFABP of other species all indicate comparable binding affinity to human LFABP (P > 0.05) for all PFAS.

The analysis of each specific energy term of $\Delta G_{bind}$ shows that for all LFABP of different species, electrostatic interaction makes the most significant contribution to $\Delta G_{bind}$ for PFAS, although most of the electrostatic interaction is compensated by the polar solvation energy. The nonpolar solvation energies are very small and remain stable among ligands, thus making minor contribution to the binding free energy (Supplemental Data, Table S2-S8). Finally, as indicated in Supplemental Data Figure S4-S7, a strong negative relationship is observed between van der Waals and carbon chain length for all LFABP of different species: as carbon chain length increases, van der Waals decreases. The entropy component also shows a similar trend, but the correlation is weaker than van der Waals.

Finally, a correlation analysis was performed between $\Delta G_{bind}$ and carbon chain length. As indicated in Figure 6, in all LFABP systems, a quite strong negative relationship was observed for both LFABP versus PFCAs and LFABP versus PFSAs, with the correlation coefficient ranging from −0.64 to −1.0.

## DISCUSSION

In this study, we demonstrated how to combine the SeqAPASS and the multi-step molecular modeling workflow to inform PFAS bioaccumulation potential across species. SeqAPASS informs PFAS bioaccumulation potential based on the analysis of protein sequence conservation. The SeqAPASS evaluation for LFABP conservation in this example provides a prediction that identifies species as likely or unlikely to have similar bioaccumulation potential to human. By assessing protein sequence similarity at different levels, SeqAPASS predicted similar PFAS bioaccumulation potential between humans and the majority of vertebrates when considering those critical amino acid residues that form hydrogen bonds

with PFAS with the exception of a subset of vertebrates that include the Japanese medaka and fathead minnow.

Further, the SeqAPASS tool can be used for hypothesis generation via the identification of differences in amino acids across taxa. Therefore, to follow-up from previous studies identifying differences among human and rat predicted binding affinities to PFAS, SeqAPASS Level 3 identified a number of potential critical individual residues (i.e., F50, A54, T81, T93, and N97) that differed between human LFABP and LFABPs in other species. Based on sidechain classification and molecular weight as a surrogate for size, those residue differences among species groups provide insights as to which residues are likely to influence interactions with PFAS and therefore change the bioaccumulation potential. Additional evidence that these individual amino acid mutations found across species may be important was generated using the DUET tool to mutate those single amino acids in the human structure to represent residues seen in other species, where destabilization of the protein was found. Therefore, to more fully understand whether these predicted amino acid differences determined based primarily on sequence translate to changes in binding affinity based on structure the multi-step molecular modeling workflow was employed. By evaluating changes in individual amino acids and double mutations at the structural level using molecular dynamics, insights were gained regarding which amino acids may or may not be important for PFAS-LFABP interaction. In considering the five amino acid differences identified by SeqAPASS, only F50V mutation lead to significant differences in binding to PFNA, showing stronger binding affinity than the wild-type human LFABP (Figure 4), therefore demonstrating the utility of the combined methods for hypothesis generation. However, the molecular modeling workflow also showed that in most cases, single or double mutation of these residues which distinguish one species to another species did not significantly change their binding affinity (Figure 4), which suggests that these residue difference solely may not significantly affect LFABP binding affinity for PFAS.

Building upon and advancing the sequence-based predictions, the multi-step molecular modeling workflow provides additional PFAS bioaccumulation potential information from the perspective of protein function (e.g., binding affinity) from evaluation of the full sequences. By estimating PFAS binding affinity to structural models of LFABP across different species (i.e., human, rat, chicken, zebrafish, rainbow trout, Japanese medaka and fathead minnow), our workflow revealed that human LFABP has comparable PFAS binding affinity to all other vertebrate species evaluated, except Japanese medaka and fathead minnow. The LFABP of those two fish species indicated significantly weaker binding affinities than human for some PFAS ligands (Table 3 and Figure 5). A closer look at the binding mode of PFAS bound to human, Japanese medaka, and fathead minnow LFABP shows that the close contact residues are very similar across those species for different PFAS, but the positions of these residues are quite different between human and the two fish species (e.g., SER124 versus SER52, Supplemental Data Table S9 and Figure S8). It seems that the position of key residues, which seem to drive the position of ligand binding, can cause significantly different binding affinities between humans and the two fish species. Because the identity, not the position, of close-contact residues is conserved (i.e., the residue is a serine in both cases in the example above), the specific amino acids are implicated in facilitating certain key interactions (e.g. hydrogen bonding). At the same time, when the

position of ligands is closer to the bottom of the LFABP binding pocket, the binding affinity also tends to be stronger (Figure S8). Thus, we conclude that when the position of key residues facilitate binding in a region of the protein that is more energetically favorable (e.g. increases hydrophobic contacts), stronger binding affinities result. However, these observations should be tempered with an acknowledgment that molecular simulations have a degree of uncertainty and variations in the predictions of exact binding conformations can and do occur from simulation to simulation.

Given the SeqAPASS and molecular modeling results, human, rat, chicken, zebrafish or rainbow trout seem to be better representative species of the higher range of vertebrate bioaccumulation potential of PFAS than Japanese medaka and fathead minnow. It is worth pointing out that this conclusion is based on the interaction between PFAS and LFABP. Other proteins such as serum albumin and membrane transporters also play important roles in determining the bioaccumulation behavior of PFAS (Cheng and Ng 2017) and should be included in future work.

In addition to offering fast and reliable estimation of protein binding affinity for PFAS across different species, which significantly expands the cross-species evaluation beyond yes or no predictions of bioaccumulation potential generated by SeqAPASS, the inclusion of molecular dynamics in this multi-step molecular modeling workflow provides valuable insights into how the chemical structures of PFAS influence their protein binding behavior. For example, the $\Delta G_{bind}$ results for different species showed that a strong negative correlation exists between the carbon chain length of PFAS and their LFABP binding affinity. A further examination of each individual energy component of $\Delta G_{bind}$ indicates that this strong correlation is largely due to the van der Waals interactions and entropy change, both of which have a close relationship with carbon chain length (Figure S4–S7 Supplemental Data). The tools evaluated in this study (SeqAPASS, DUET, and molecular dynamics) are highly complementary, with unique strengths and weaknesses. The major strength of SeqAPASS lies in its ability to rapidly predict bioaccumulation potential for hundreds of species and further screen for potentially important differences across hundreds of proteins. Its main limitation is that the type of difference observed across sequences may not necessarily translate to a susceptibility difference based on structural details of a protein/receptor (e.g., whether that difference lies within a ligand-binding domain or results in a large enough structural shift). This weakness can be overcome by coupling to a structure-based tool such as molecular dynamics. While requiring considerably more time and computational resources than SeqAPASS, molecular dynamics provides valuable structural insight into key differences across species. Finally, the differences observed in our single-residue mutations compared with whole-species sequences indicates that the additional structural information provided by comparing "wild type" proteins among species can be valuable, particularly in determining whether a "model species" is indeed a good model for predicting bioaccumulation potential or toxicity in humans.

Finally, it is worth pointing out that beyond this investigation of PFAS-LFABP interaction, the combined SeqAPASS and molecular modeling workflow can be applied to other ligand-protein pairs to provide insights into biological distribution, tissue-specific bioaccumulation driven by protein binding, and/or toxicity based on specific receptor interactions for any

chemical and biological target of interest. For example, studies have shown PFAS could cause developmental and immunotoxic effects via pathways that may depend on their interaction with peroxisome proliferator-activated receptors (PPARs) or may be independent of these receptors, and this dependence may differ by species (e.g., humans vs rodents) (Abbott et al. 2012; DeWitt et al. 2009; Szilagyi et al. 2020). Our workflow can then be employed to examine the interactions between PPARs and PFAS in different species to understand relative strengths of interaction. To generalize the workflow to other ligand-protein pairs, only the protein sequences and the 3-dimensional crystal structures for the proteins and ligands are required as input data. Moreover, it is useful to have some experimental data (e.g., protein binding affinity for ligands) for model evaluation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

## REFERENCES

Abbott BD, Wood CR, Watkins AM, Tatum-Gibbs K, Das KP, Lau C (2012). Effects of perfluorooctanoic acid (pfoa) on expression of peroxisome proliferator-activated receptors (ppar) and nuclear receptor-regulated genes in fetal and postnatal cd-1 mouse tissues. Reprod. Toxicol 33(4), 491–505. [PubMed: 22154759]

Burkhard LP (2020). (In review currently but will be published prior to this manuscript). Evaluation of published bioaccumulation data for per- and polyfluoroalkyl substances across aquatic species.

Caldwell GW, Yan Z 2004. Isothermal titration calorimetry characterization of drug-binding energetics to blood proteins. Methods Pharmacol. Toxicol 123–149.

Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2014). AMBER 14. University of California, San Francisco.

Cheng W, Ng CA (2017). A permeability-limited physiologically based pharmacokinetic (pbpk) model for perfluorooctanoic acid (PFOA) in male rats. Environ. Sci. Technol 51(17), 9930–9939. [PubMed: 28759222]

Cheng W, Ng CA (2018). Predicting relative protein affinity of novel per-and polyfluoroalkyl substances (PFASs) by an efficient molecular dynamics approach. Environ. Sci. Technol 52(14), 7972–7980. [PubMed: 29897239]

DeWitt JC, Shnyra A, Badr MZ, Loveless SE, Hoban D, Frame SR, Cunard R, Anderson SE, Meade BJ, Peden-Adams MM (2009). Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha. Crit. Rev. Toxicol 39(1), 76–94. [PubMed: 18802816]

Doering JA, Lee S, Kristiansen K, Evenseth L, Barron MG, Sylte I, LaLone CA (2018). In silico site-directed mutagenesis informs species-specific predictions of chemical susceptibility derived from the sequence alignment to predict across species susceptibility (SeqAPASS) tool. Toxicol. Sci 166(1), 131–145. [PubMed: 30060110]

Han X, Snow TA, Kemper RA, Jepson GW (2003). Binding of perfluorooctanoic acid to rat and human plasma proteins. Chem. Res. Toxicol 16(6), 775–781. [PubMed: 12807361]

Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. J Cheminf. 4(1), 17.

Homeyer N, Gohlke H (2012). Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Mol. Inf 31(2), 114–122.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys 79(2), 926–935.

Kastritis PL, Bonvin AM (2013). On the binding affinity of macromolecular interactions: Daring to ask why proteins interact. J. R. Soc. Interface 10(79), 20120835. [PubMed: 23235262]

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015). The phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc 10(6), 845. [PubMed: 25950237]

Kennedy GL, Butenhoff JL, Olsen GW, O'Connor JC, Seacat AM, Perkins RG, Biegel LB, Murphy SR, Farrar DG (2004). The toxicology of perfluorooctanoate. Crit. Rev. Toxicol 34(4), 351–384. [PubMed: 15328768]

Kim S-J, Heo S-H, Lee D-S, Hwang IG, Lee Y-B, Cho H-Y (2016). Gender differences in pharmacokinetics and tissue distribution of 3 perfluoroalkyl and polyfluoroalkyl substances in rats. Food Chem. Toxicol 97, 243–255. [PubMed: 27637925]

Kudo N, Kawashima Y (2003). Toxicity and toxicokinetics of perfluorooctanoic acid in humans and animals. J. Toxicol. Sci 28(2), 49–57. [PubMed: 12820537]

LaLone CA, Villeneuve DL, Lyons D, Helgen HW, Robinson SL, Swintek JA, Saari TW, Ankley GT (2016). Editor's highlight: Sequence alignment to predict across species susceptibility (SeqAPASS): A web-based tool for addressing the challenges of cross-species extrapolation of chemical toxicity. Toxicol Sci. 153(2), 228–245. [PubMed: 27370413]

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. 43(D1), D222–D226. [PubMed: 25414356]

Miller BR III, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012). MMPBSA.py: an efficient program for end-state free energy calculations. J. Chem. Theory Comput 8(9), 3314–3321. [PubMed: 26605738]

Morris GM, Lim-Wilby M (2008). Molecular docking methods. Mol. Biol 443, 365–382.

Ng CA, Hungerbühler K (2013). Bioconcentration of perfluorinated alkyl acids: How important is specific binding? Environ. Sci. Technol 47(13), 7214–7223. [PubMed: 23734664]

Ng CA, Hungerbühler K (2014). Bioaccumulation of perfluorinated alkyl acids: Observations and models. Environ. Sci. Technol 48(9), 4637–4648. [PubMed: 24762048]

Olsen GW, Burris JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, Zobel LR (2007). Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. Environ. Health Perspect 115(9), 1298–1305. [PubMed: 17805419]

Pires DE, Ascher DB, Blundell TL (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. 42(W1), W314–W319. [PubMed: 24829462]

Sheng N, Cui R, Wang J, Guo Y, Wang J, Dai J (2018). Cytotoxicity of novel fluorinated alternatives to long-chain perfluoroalkyl substances to human liver cell line and their binding capacity to human liver fatty acid binding protein. Arch. Toxicol 92(1), 359–369. [PubMed: 28864880]

Sheng N, Li J, Liu H, Zhang A, Dai J (2016). Interaction of perfluoroalkyl acids with human liver fatty acid-binding protein. Arch. Toxicol 90(1), 217–227. [PubMed: 25370009]

Szilagyi JT, Avula V, Fry RC (2020). Perfluoroalkyl substances (PFAS) and their effects on the placenta, pregnancy, and child development: a potential mechanistic role for placental peroxisome proliferator–activated receptors (PPARs). Curr. Envir. Health Rpt 7, 222–230.

Trott O, Olson AJ (2010). Autodock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem 31(2), 455–461. [PubMed: 19499576]

Wang Z, DeWitt JC, Higgins CP, Cousins IT (2017). A never-ending story of per-and polyfluoroalkyl substances (PFASs)? Environ. Sci. Technol 51(5), 2508–2518. [PubMed: 28224793]

Weaver YM, Ehresman DJ, Butenhoff JL, Hagenbuch B (2009). Roles of rat renal organic anion transporters in transporting perfluorinated carboxylates with different chain lengths. Toxicol. Sci 113(2), 305–314. [PubMed: 19915082]

Woodcroft MW, Ellis DA, Rafferty SP, Burns DC, March RE, Stock NL, Trumpour KS, Yee J, Munro K (2010). Experimental characterization of the mechanism of perfluorocarboxylic acids' liver protein bioaccumulation: The key role of the neutral species. Environ. Toxicol. Chem 29(8), 1669–1677. [PubMed: 20821618]

Yang C-H, Glover KP, Han X (2009). Organic anion transporting polypeptide (oatp) 1a1-mediated perfluorooctanoate transport and evidence for a renal reabsorption mechanism of oatp1a1 in renal elimination of perfluorocarboxylates in rats. Toxicol. Lett 190(2), 163–171. [PubMed: 19616083]

Yang C-H, Glover KP, Han X (2010). Characterization of cellular uptake of perfluorooctanoate via organic anion-transporting polypeptide 1a2, organic anion transporter 4, and urate transporter 1 for their potential roles in mediating human renal reabsorption of perfluorocarboxylates. Toxicol. Sci 117(2), 294–302. [PubMed: 20639259]

$$\Delta G_{bind} = \frac{1}{N} \sum_{i=1}^{N} G_i^{Complex} - G_i^{Protein} - G_i^{PFAS}$$

$$G = E_{bond} + E_{electrostatic} + E_{van\ der\ Waals} + G_{polar} + G_{nonpolar} - TS$$
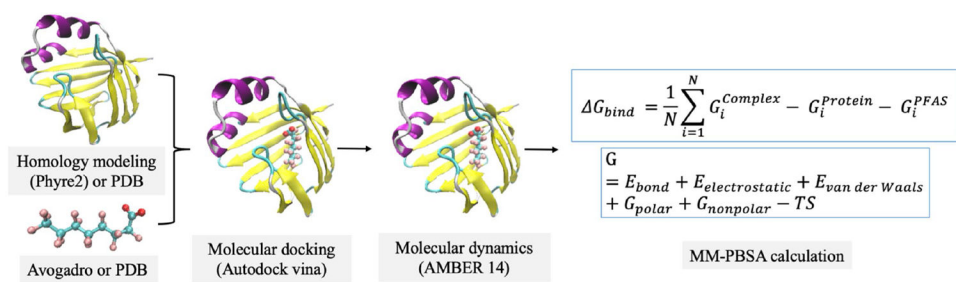
**Figure 1.**
Multi-step Molecular Modeling Workflow
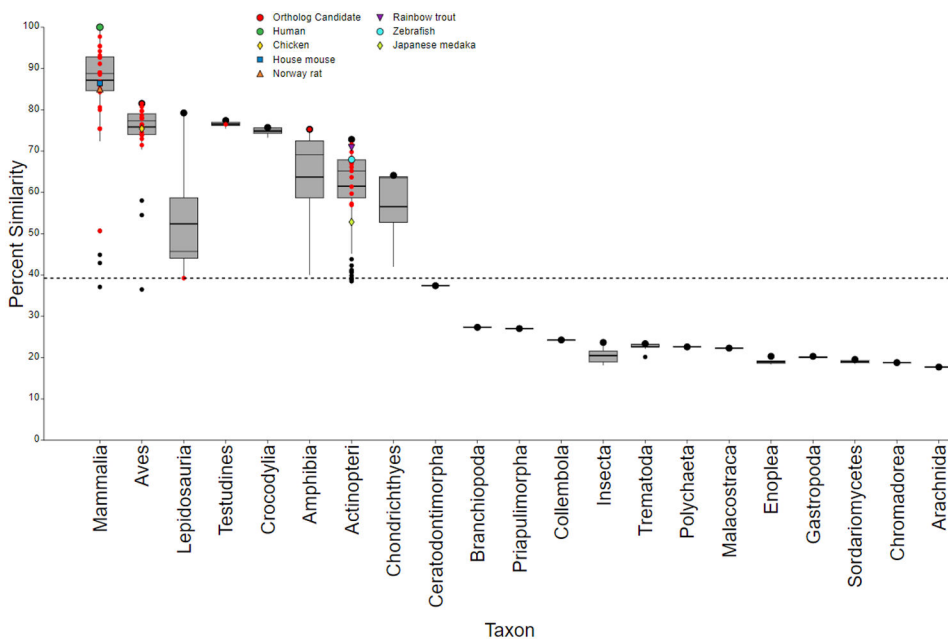
**Figure 2.**
SeqAPASS (v4.0) results illustrating the percent similarity across species compared to human (Homo sapiens) liver fatty acid binding protein (LFABP) aligning primary amino acid sequences. The green open circle (○) represents the human LFABP and black filled circle (O) represents the species with the highest percent similarity within the specified taxonomic group. Red dots represent ortholog candidates. The top and bottom of each box represent the 75th and 25th percentiles, respectively. The top and bottom whiskers extend up to 1.5 times the interquartile range. The mean and median values for each taxonomic group are represented by horizontal thick and thin black lines on the box, respectively. The dashed line indicates the cut-off for predictions of LFABP conservation with those taxonomic groups above or crossing through the cut-off predicted to have similar bioaccumulation potential for PFAS as human and those below likely to be different.
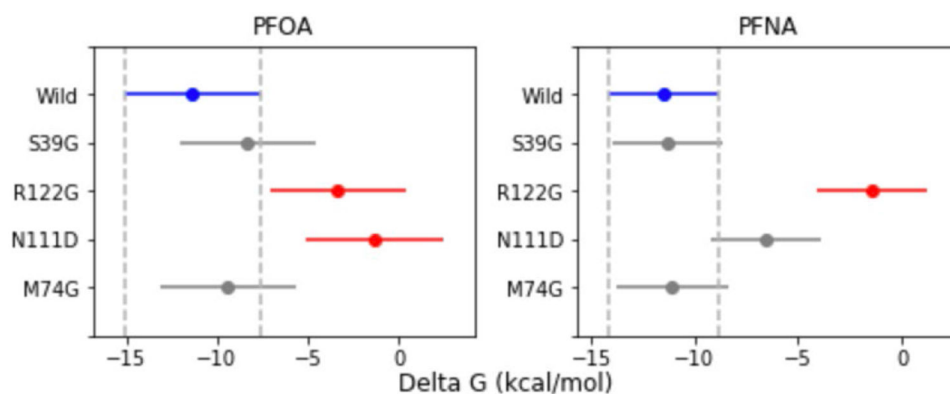
**Figure 3.**
Multiple comparison (Tukey test) between wild-type human LFABP and single mutated LFABP mean free energy (Delta G) measures for PFOA and PFNA. Mutations (e.g., S39G, meaning residue at position 39 was mutated from S to G, similar for other mutations) were selected for comparison with experimental data from the Sheng et al. study. Blue is wild-type human LFABP; red indicates significant difference ($p < 0.05$); gray means no significant difference from wild-type ($p > 0.05$). The error bars represent 95% confidence interval for each dataset.
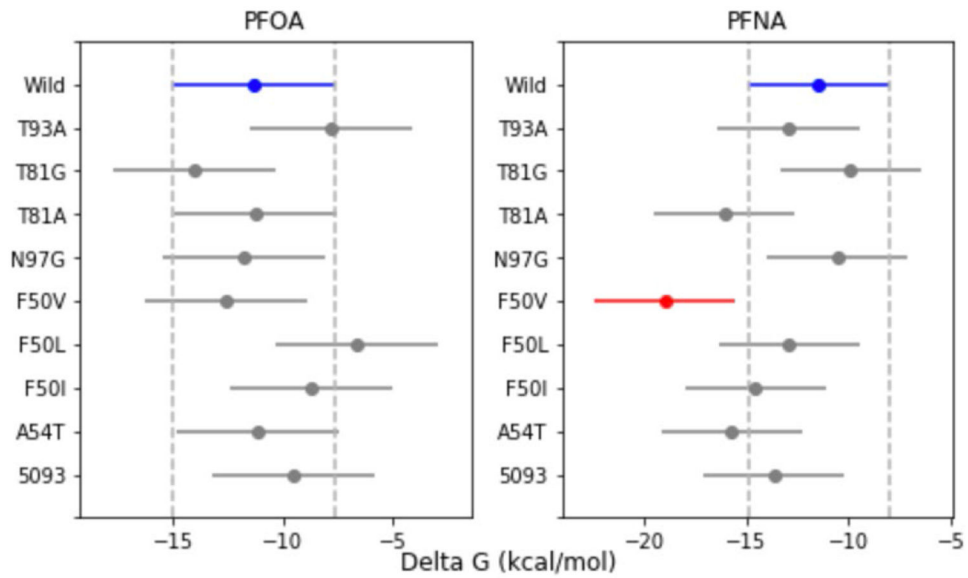
**Figure 4.**
Multiple comparison between wild-type human LFABP and mutated LFABP (Tukey test) for
PFOA and PFNA, those mutations (T93F means residue at position 93 mutated from T to F,
similar for other mutations) were selected based on SeqAPASS Level 3 results. Blue is wild-
type human LFABP; red indicates significant difference ($p < 0.05$); gray means no
significant difference ($p > 0.05$). The error bars represent 95% confidence interval for each
dataset.
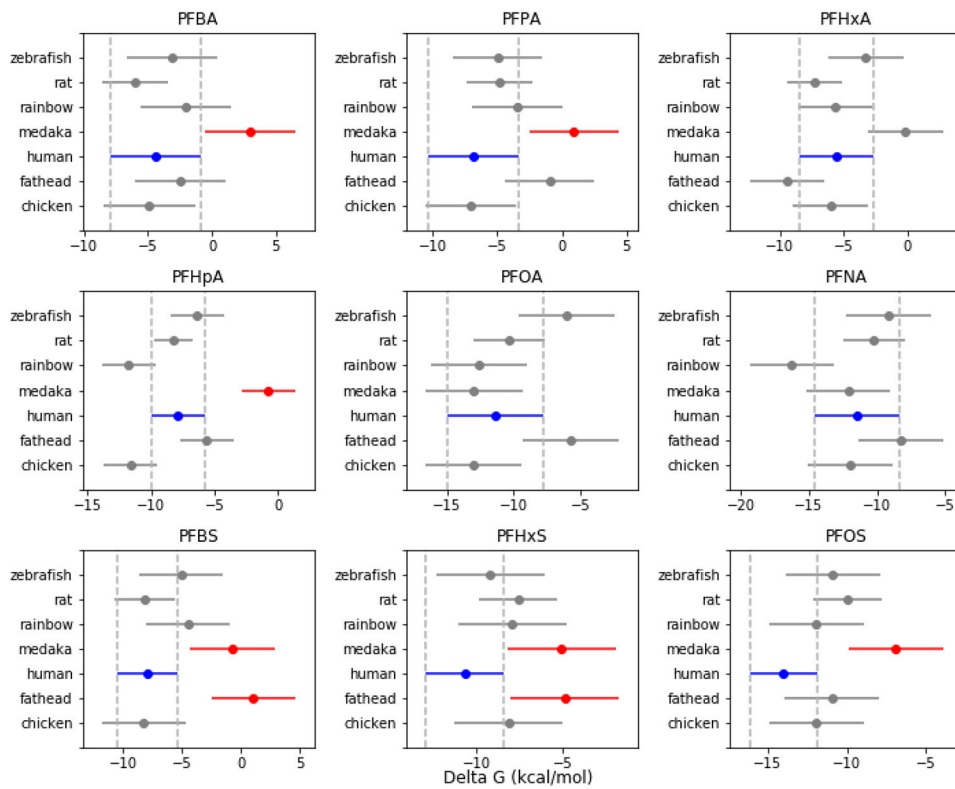
**Figure 5.**
Multiple comparison (Tukey test) between human LFABP and other LFABPs for different PFAS. Blue is human LFABP; red indicates significant difference (p < 0.05); gray means no difference from human wild type (p > 0.05). The error bars represent 95% confidence interval for each dataset.
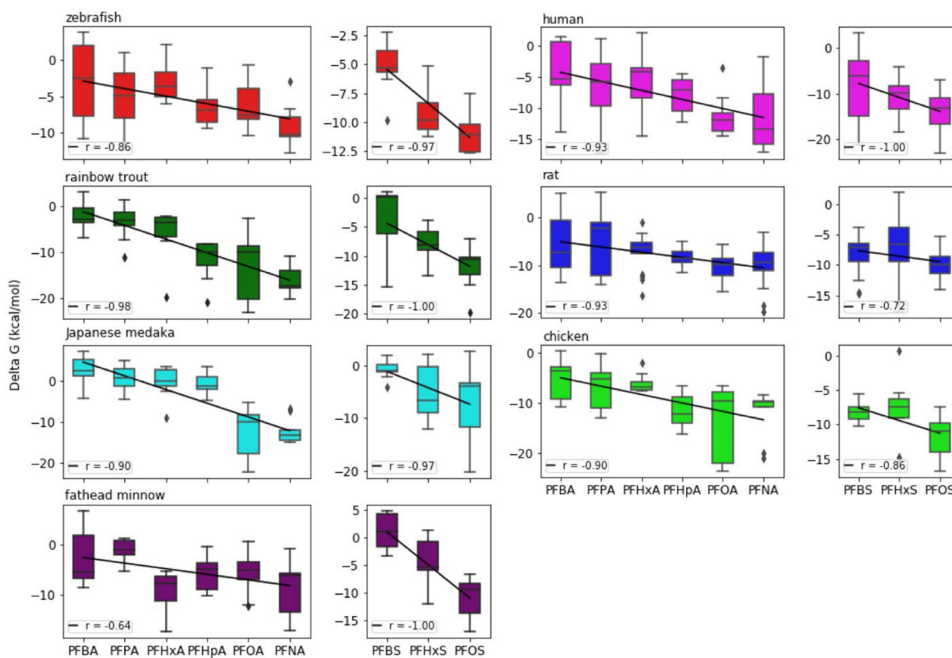
**Figure 6.**
Distribution of $\Delta G_{bind}$ for different PFAS-LFABP complexes across species. The correlation coefficient between the mean of $\Delta G_{bind}$ and PFAS chain length ranges between −0.64 and −1.00, indicating that binding affinity increases with chain length (a more negative $\Delta G_{bind}$ means stronger affinity, and is exponentially related to the dissociation constant, $K_D$).

**Table 1.**

Structural information for the 9 PFAS included in this study
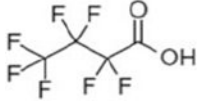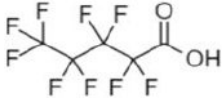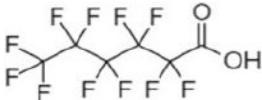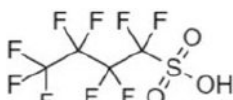
| Name | Acronym | CAS number | Carbon chain length | 2D structures |
|------|---------|------------|---------------------|---------------|
| perfluorobutanoic acid | PFBA | 375-22-4 | 4 | |
| perfluoropentanoic acid | PFPA | 2706-90-3 | 5 | |
| perfluorohexanoic acid | PFHxA | 307-24-4 | 6 | |
| perfluoroheptanoic acid | PFHpA | 375-85-9 | 7 | |
| perfluorooctanoic acid | PFOA | 335-67-1 | 8 | |
| perfluorononanoic acid | PFNA | 375-95-1 | 9 | |
| perfluorobutane sulfonate | PFBS | 375-73-5 | 4 | |
| perfluorohexane sulfonate | PFHxS | 355-46-4 | 6 | |
| perfluorooctane sulfonate | PFOS | 1763-23-1 | 8 | |

**Table 2.**

Amino acid differences across species compared to human LFABP that predict different bioaccumulation potential.

| Human Amino Acid Position | Type 1 Primates, Ruminants, Whales/dolphins | Type 2 Rodents and other mammals, Fish, Amphibians, Testudines | Type 3 Aves, Lepidosauria Chondrichthyes | Type 4 Crocodylia | SeqAPASS Level 3 Prediction of Similar to Human LFABP Template | Mutation in DUET | Stability Change from DUET ($\Delta G$, kcal/mol) |
|---|---|---|---|---|---|---|---|
| 50 | Phenylalanine (F) | Valine (V) | Valine | Phenylalanine | Yes | F50V | −1.196 (Destabilizing) |
|  |  | Isoleucine (I) | Isoleucine |  | No | F50I | −0.808 (Destabilizing) |
|  |  | Leucine (L) | Leucine |  | No | F50L | −0.893 (Destabilizing) |
| 54 | Alanine (A) | Threonine (T) | Threonine | Threonine | Yes | A54T | −0.195 (Destabilizing) |
|  |  |  |  |  | No |  |  |
| 81 | Threonine | Alanine | Alanine | Threonine | Yes | T81A | −0.749 (Destabilizing) |
|  |  | Glycine (G) |  |  | No | T81G | −0.023 (Destabilizing) |
|  |  |  |  |  | No |  |  |
| 93 | Threonine | Threonine | Alanine |  | Yes | T93V | 0.031 (Stabilizing) |
|  |  | Valine | Alanine |  | Yes | T93A | −1.004 (Destabilizing) |
|  |  |  |  |  | No |  |  |
| 97 | Asparagine (N) | Glycine | Glycine | Glycine | Yes | N97G | 0.521 (Stabilizing) |
|  |  |  |  |  | No |  |  |

**Table 3.**

The average, maximum, and minimum predicted free energy of binding ( $G_{bind}$, kcal/mol) over all tested PFAS ligands for 7 different species of LFABP.

| LFABPs | Max $G_{bind}$ | Min $G_{bind}$ | Mean $G_{bind}$ |
|---|---|---|---|
| human | −4.39333 | −13.9894 | −8.89 |
| rat | −4.85333 | −10.3439 | −8.06698 |
| chicken | −4.89333 | −12.9956 | −9.2 |
| zebrafish | −3.12778 | −10.8956 | −6.44444 |
| rainbow trout | −2.01111 | −16.2389 | −8.45975 |
| Japanese medaka | 2.956667 | −12.9867 | −3.86617 |
| fathead minnow | 1.024444 | −10.9344 | −5.25457 |

**Table 4.**

ANOVA test between 7 different species of LFABP (human, rat, chicken, and zebrafish, rainbow trout, Japanese medaka and fathead minnow) for different ligands.

| Ligands | F value | P value |
|---------|---------|---------|
| PFBA | 3.690712 | 0.003231 |
| PFPA | 3.43111 | 0.005272 |
| PFHxA | 5.015558 | 0.000281 |
| PFHpA | 15.06468 | 1.02E-10 |
| PFOA | 3.557624 | 0.004152 |
| PFNA | 3.350566 | 0.006139 |
| PFBS | 6.07476 | 3.29E-05 |
| PFHxS | 2.682188 | 0.020687 |
| PFOS | 3.285783 | 0.006392 |