Supplementary Material for:

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*], and 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
[*]Co-corresponding authors
[4]Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA

There are **19** pages, **8** figures and **9** tables included in the supplementary material.

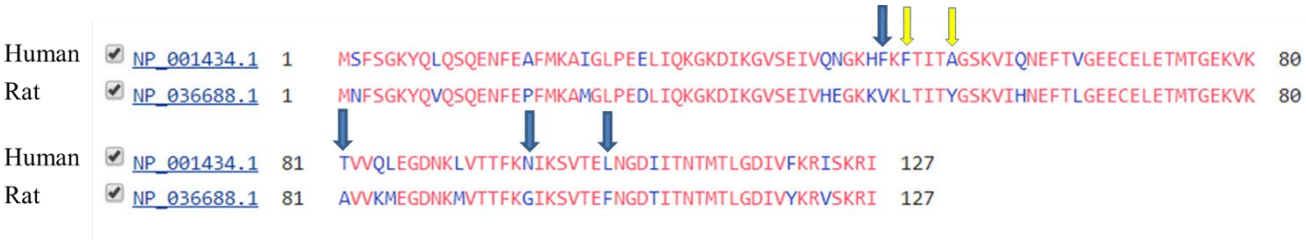

Figure S1. Human LFABP aligned with Norway Rat LFABP. Blue text represents sequence differences in individual amino acids. Arrows indicate substitutions that would be predicted by SeqAPASS Level to lead to differences in chemical binding to LFABP based on differences in side chain classification and molecular weight differences in amino acids greater than 30 g/mol (Doering et al. 2018). Yellow arrows indicate the position of the amino acids where substitutions were most commonly identified across species.

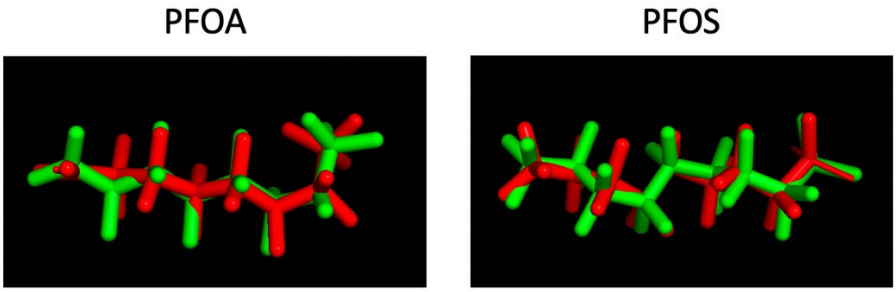

Figure S2. Comparison of 3-dimensional structures between experimentally determined structures (obtained from the Protein Data Bank, red color) and structures drawn in Avogadro (green color) for PFOA and PFOS. Structure alignment was generated with PyMOL (Schrodinger 2015).

Human:

>3STM_1|Chain X|Fatty acid-binding protein, liver|Homo sapiens (9606)
MAGSSFSGKYQLQSQENFEAFMKAIGLPEELIQKGKDIKGVSEIVQNGKHFKFTITAGSKVIQNEFTVGEECE
LETMTGEKVKTVVQLEGDNKLVTTFKNIKSVTELNGDIITNTMTLGDIVFKRISKRIGT

Rat:

>1LFO_1|Chain A|LIVER FATTY ACID BINDING PROTEIN|Rattus norvegicus (10116)
XMNFSGKYQVQSQENFEPFMKAMGLPEDLIQKGKDIKGVSEIVHEGKKVKLTITYGSKVIHNEFTLGEECELE
TMTGEKVKAVVKMEGDNKMVTTFKGIKSVTEFNGDTITNTMTLGDIVYKRVSKRI

Chicken:

>AAK58095.1 liver fatty acid binding protein [Gallus gallus]
MSFTGKYELQSHENFEPFMKALGLPDDQIQKGKDIKSISEIVQNGNKFKITVTTGSKVMTNEFTIGEECE
MELLTGEKAKCIVNMEGNNKLVANLKGLKSVTELNGDTITHTMTKGDLTYKRISKRI

Zebrafish:

>NP_001038177.1 fatty acid binding protein 1-A, liver [Danio rerio]
MAFTGKYQLESHENFEAFMKAVGVPDDEVEKGKDIKSISEIHQDGKDFKVTVTAGTKVILYSFTVGEECE
LETFTGDRAKTVVQMDGNKLTAFVKGIESVTELDGDTISNTLSFNGIVYKRISRRIS

Rainbow trout:

>XP_021446645.1 fatty acid-binding protein, liver [Oncorhynchus mykiss]
MAFTGKYQLESQENFEPFMKAIGLPDDLIQKGKDIKSVSEIEQNGDHFKVTVTTGTKVMVNSFTVGQEAE
LETLTGEKIKSTVNLVGNKLMVSLKGIESVTEFNGDTIIATMMLGPIVYKRISKRI

Japanese medaka:

>XP_004078356.1 fatty acid-binding protein, liver [Oryzias latipes]
MDFNGTWQVYSQENYEEFLRALELSEDIIKLAKDVKPVTEIKQTGNDFVITSKTPGRTVTNSFTIGKEAE
ISTMDGKKLKCVVNMEGGKLVCKTGKFCHVQEIKGGEMVETMTVGSTTLIRKSKKM

Fathead minnow:

>ABG73404.1 liver fatty acid binding protein 10, partial [Pimephales promelas]
VYLQENYEEFLPAIPLPEDIIKLAKDVKPVTEIQQKGNDFTITSKTPGKTVTNSFTVGKEAEITTMDGKK
LKCIVKLEGGKLVCNTERFSHIQEIKGGEMVETLTVAGTTMVRKSKKI

Figure S3. Amino acid sequences for different species. The sequences can be downloaded from PDB website (https://www.rcsb.org/) for human and rat and from NCBI website (https://www.ncbi.nlm.nih.gov/) for other species.

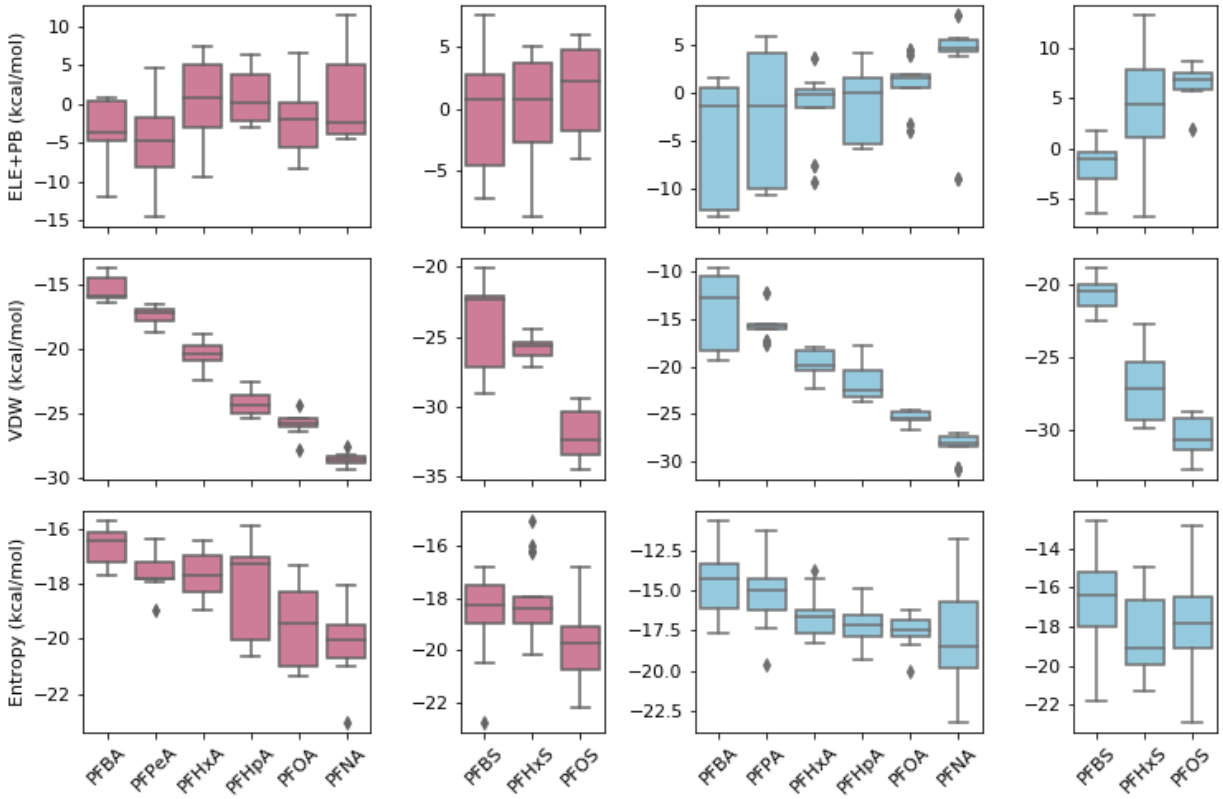

Figure S4. The distributions of the energy components of ΔG_{bind} including the sum of electrostatic interactions and polar solvation free energy (ELE + PB), van der Waals energy (vdw), and entropy change upon binding. Pink represents the human LFABP system, while blue is the rat LFABP system.

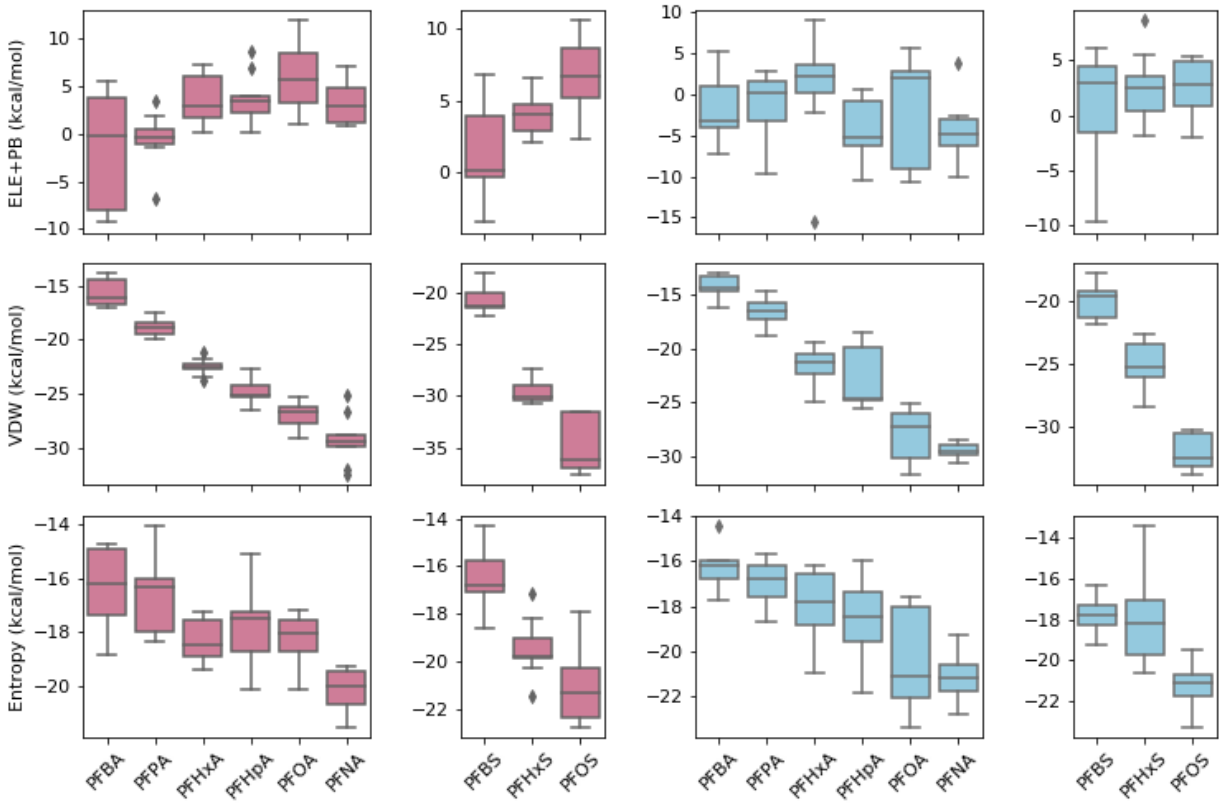

Figure S5. The distributions of the energy components of ΔG_{bind} including the sum of electrostatic interactions and polar solvation free energy (ELE + PB), van der Waals energy (vdw), and entropy change upon binding. Pink represents the zebrafish LFABP system, while blue is the rainbow trout LFABP system.

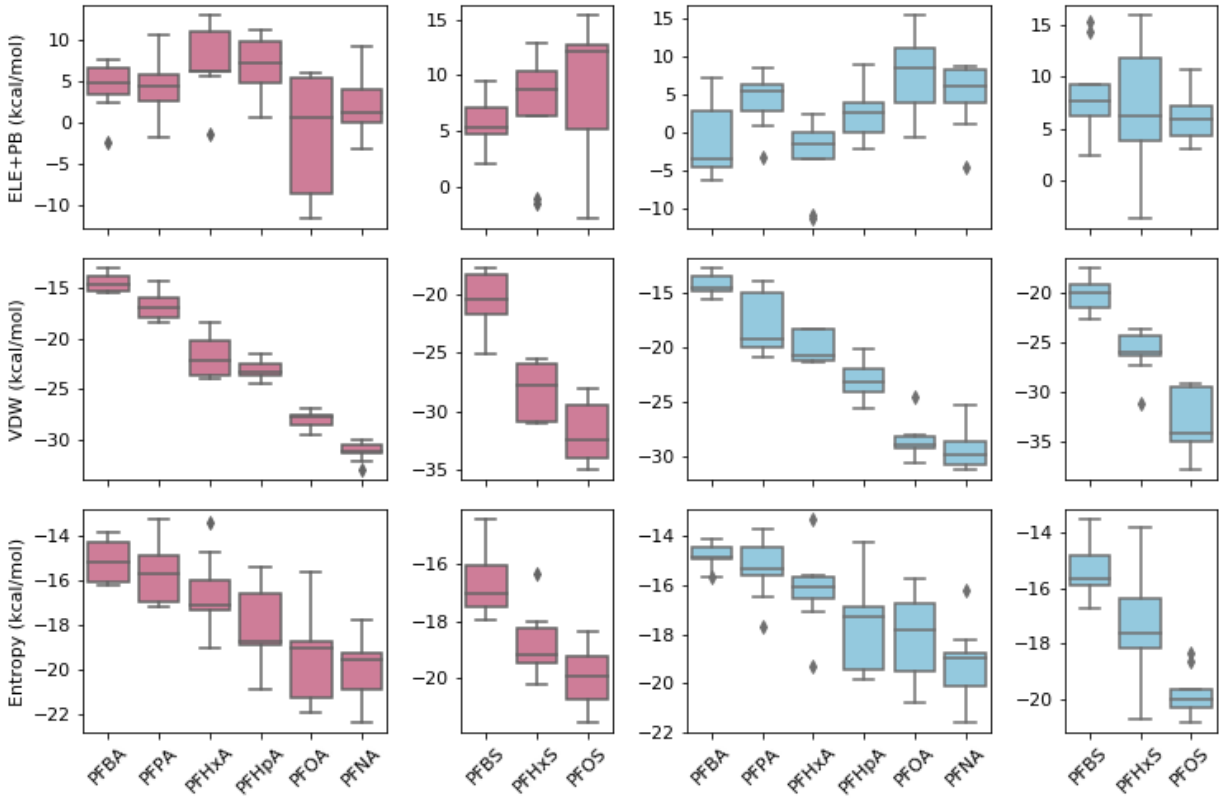

Figure S6. The distributions of the energy components of ΔG_{bind} including the sum of electrostatic interactions and polar solvation free energy (ELE + PB), van der Waals energy (vdw), and entropy change upon binding. Pink represents the Japanese medaka LFABP system, while blue is the fathead minnow LFABP system.

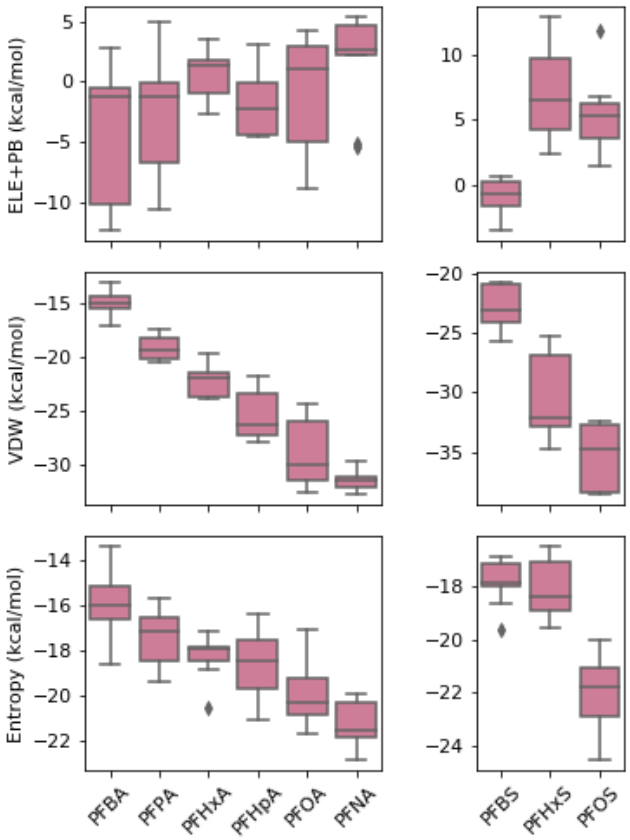

Figure S7. The distributions of the energy components of ΔG_{bind} including the sum of electrostatic interactions and polar solvation free energy (ELE + PB), van der Waals energy (vdw), and entropy change upon binding for chicken LFABP system.

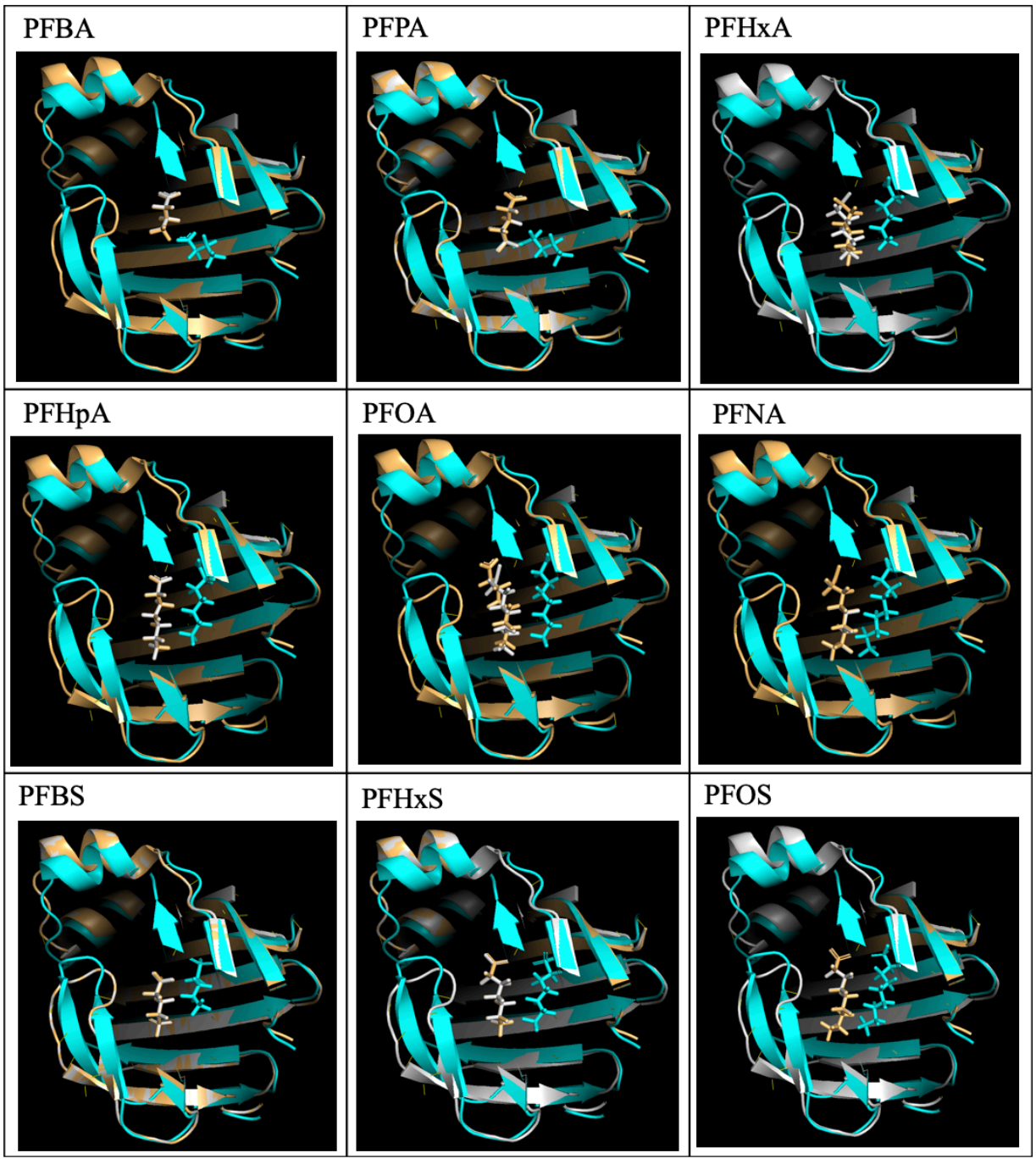

Figure S8. The PFAS binding poses for human (cyan color), Japanese medaka (orange color) and fathead minnow (grey color) LFABP after sequence alignment. The results were generated by Autodock vina (Trott and Olson 2010) and visualized with PyMOL (Schrodinger 2015)

Table S1. Links to EPA's CompTox Chemicals Dashboard for all PFAS.

PFAS	Link to Dashboard
PFBA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4059916
PFPA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8059970
PFHxA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID3031862
PFHpA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID1037303
PFOA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8031865
PFNA	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8031863
PFBS	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID5030030
PFHxS	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID80873012
PFOS	https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID3031864

Table S2. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on MM-PBSA for human LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-15.337	0.315	-78.824	6.356	75.159	5.391	-2.003	0.013	-16.616	0.203	-4.393	1.545
PFPA	-17.460	0.218	-99.408	4.396	94.660	2.832	-2.253	0.015	-17.609	0.226	-6.851	1.773
PFHxA	-20.558	0.334	-78.300	7.409	78.159	5.698	-2.586	0.020	-17.730	0.296	-5.553	1.651
PFHpA	-24.340	0.295	-69.928	3.429	71.008	3.918	-2.828	0.026	-18.214	0.574	-7.872	0.953
PFOA	-25.872	0.289	-98.968	3.093	97.117	1.977	-3.187	0.017	-19.564	0.509	-11.344	1.100
PFNA	-28.597	0.170	-95.384	3.393	95.901	1.800	-3.510	0.017	-20.133	0.436	-11.454	1.724
PFBS	-23.959	0.740	-90.748	4.010	90.734	3.459	-2.436	0.014	-18.507	0.328	-7.901	1.875
PFHxS	-25.756	0.196	-88.319	2.010	88.321	1.226	-3.016	0.009	-18.117	0.311	-10.651	0.905
PFOS	-31.966	0.418	-88.897	2.414	90.629	1.702	-3.578	0.015	-19.821	0.316	-13.989	1.102

Table S3. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for rat LFABP (All values have units of kcal/mol)

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-13.870	0.854	-98.032	5.124	93.354	3.998	-1.859	0.036	-14.387	0.473	-6.019	1.347
PFPA	-15.723	0.341	-95.694	5.151	93.729	4.024	-2.232	0.042	-15.067	0.484	-4.853	1.493
PFHxA	-19.640	0.322	-100.333	2.888	98.757	2.202	-2.608	0.011	-16.584	0.303	-7.241	0.877
PFHpA	-21.700	0.467	-102.699	2.352	101.938	1.536	-2.937	0.017	-17.191	0.313	-8.206	0.443
PFOA	-25.480	0.160	-105.327	1.324	106.190	0.973	-3.208	0.008	-17.481	0.206	-10.344	0.619
PFNA	-28.430	0.315	-104.583	3.227	108.249	2.335	-3.496	0.011	-18.032	0.637	-10.226	0.961
PFBS	-20.629	0.255	-101.672	1.512	99.922	1.112	-2.464	0.007	-16.659	0.511	-8.184	0.721
PFHxS	-26.840	0.568	-101.733	2.094	105.687	1.217	-3.072	0.017	-18.389	0.438	-7.572	1.202
PFOS	-30.561	0.312	-92.358	1.409	98.837	1.193	-3.636	0.012	-17.759	0.539	-9.959	0.531

Table S4. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for chicken LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-15.080	0.370	-89.077	6.489	85.431	4.802	-2.083	0.015	-15.911	0.461	-4.893	1.310
PFPA	-19.259	0.335	-89.921	2.681	86.956	2.075	-2.317	0.028	-17.488	0.388	-7.054	1.450
PFHxA	-22.217	0.436	-87.180	2.664	87.671	2.362	-2.576	0.021	-18.258	0.317	-6.044	0.578
PFHpA	-25.641	0.719	-101.372	2.468	99.681	2.123	-2.886	0.014	-18.628	0.510	-11.589	0.972
PFOA	-28.928	1.014	-99.812	3.446	98.859	1.946	-3.100	0.034	-19.986	0.460	-12.996	2.292
PFNA	-31.472	0.295	-96.261	3.758	98.007	2.687	-3.479	0.034	-21.250	0.298	-11.953	1.528
PFBS	-22.807	0.638	-81.700	2.544	80.894	2.543	-2.459	0.020	-17.850	0.275	-8.221	0.462
PFHxS	-30.382	1.097	-81.564	2.668	88.767	2.557	-3.020	0.022	-18.067	0.339	-8.136	1.503
PFOS	-35.583	0.862	-94.822	1.821	100.076	1.477	-3.589	0.012	-21.998	0.454	-11.913	0.943

Table S5. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for zebrafish LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-15.750	0.402	28.159	7.956	-29.751	6.489	-2.069	0.020	-16.284	0.463	-3.128	1.742
PFPA	-18.850	0.251	39.594	4.171	-40.091	3.568	-2.316	0.018	-16.676	0.464	-4.984	1.323
PFHxA	-22.507	0.249	38.133	2.499	-34.534	2.152	-2.649	0.018	-18.302	0.258	-3.253	0.805
PFHpA	-24.731	0.331	33.366	2.102	-29.816	1.595	-2.977	0.015	-17.821	0.469	-6.340	0.855
PFOA	-26.941	0.361	30.680	2.718	-24.814	2.070	-3.242	0.029	-18.299	0.328	-6.021	0.967
PFNA	-29.276	0.716	34.033	2.961	-30.606	2.845	-3.508	0.013	-20.203	0.244	-9.152	0.936
PFBS	-20.687	0.461	31.353	2.510	-29.753	1.544	-2.547	0.008	-16.583	0.388	-5.047	0.707
PFHxS	-29.532	0.368	30.416	1.212	-26.363	1.306	-3.084	0.020	-19.389	0.387	-9.179	0.696
PFOS	-34.882	0.830	29.460	2.971	-22.782	2.290	-3.593	0.023	-20.904	0.499	-10.896	0.578

Table S6. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for rainbow trout LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-14.276	0.327	-57.766	4.371	55.807	3.207	-2.099	0.014	-16.322	0.294	-2.011	0.949
PFPA	-16.662	0.461	-48.192	4.536	46.824	3.286	-2.364	0.011	-16.939	0.335	-3.457	1.185
PFHxA	-21.787	0.619	-56.374	6.261	57.190	4.129	-2.629	0.013	-17.908	0.498	-5.692	1.782
PFHpA	-22.977	0.910	-67.642	2.401	63.061	3.201	-2.944	0.024	-18.744	0.553	-11.757	1.346
PFOA	-27.943	0.803	-66.754	1.401	64.893	1.554	-3.188	0.031	-20.349	0.707	-12.648	2.337
PFNA	-29.450	0.242	-72.866	2.151	68.380	1.630	-3.549	0.019	-21.242	0.330	-16.239	0.983
PFBS	-19.956	0.462	-51.694	5.080	51.876	3.435	-2.524	0.020	-17.808	0.268	-4.489	2.107
PFHxS	-25.224	0.621	-53.624	2.378	56.066	1.878	-3.046	0.031	-17.901	0.709	-7.929	0.869
PFOS	-32.020	0.447	-61.322	3.479	63.863	3.025	-3.712	0.010	-21.273	0.333	-11.917	1.174

Table S7. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for Japanese medaka LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-14.443	0.306	-96.134	4.624	100.474	4.249	-2.077	0.017	-15.138	0.306	2.957	1.102
PFPA	-16.827	0.404	-109.589	7.659	113.956	6.520	-2.357	0.017	-15.743	0.443	0.928	1.044
PFHxA	-21.659	0.628	-84.570	4.983	91.969	4.113	-2.604	0.025	-16.723	0.549	-0.146	1.231
PFHpA	-23.079	0.310	-90.961	5.366	97.976	4.458	-2.891	0.028	-18.227	0.603	-0.726	0.841
PFOA	-28.012	0.288	-120.780	6.687	119.353	4.557	-3.187	0.017	-19.640	0.647	-12.987	2.029
PFNA	-31.103	0.301	-115.243	1.287	117.661	1.802	-3.443	0.014	-20.020	0.455	-12.111	0.986
PFBS	-20.567	0.776	-105.128	3.250	110.857	2.954	-2.520	0.017	-16.629	0.363	-0.728	0.557
PFHxS	-28.030	0.748	-108.166	4.439	115.392	3.395	-3.087	0.021	-18.821	0.371	-5.070	1.657
PFOS	-31.810	0.820	-92.241	5.492	100.887	3.542	-3.652	0.018	-19.901	0.347	-6.913	2.475

Table S8. The mean and standard error (sterr) of ΔG_{bind} and its five energy components calculated based on multi-step molecular modeling for fathead minnow LFABP (All values have units of kcal/mol).

Ligand	van der Waals		Electrostatic		EPB		Nonpolar		Entropy		Delta G	
	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr	mean	sterr
PFBA	-14.316	0.301	-91.722	5.484	90.793	4.150	-2.084	0.015	-14.833	0.177	-2.496	1.749
PFPA	-18.124	0.879	-81.094	2.779	85.300	1.829	-2.277	0.011	-15.249	0.395	-0.948	0.678
PFHxA	-20.119	0.430	-108.041	3.931	105.106	2.586	-2.586	0.021	-16.180	0.492	-9.459	1.379
PFHpA	-23.071	0.538	-107.604	7.773	110.399	6.784	-2.910	0.018	-17.564	0.579	-5.622	1.021
PFOA	-28.539	0.518	-87.569	5.625	95.450	4.402	-3.146	0.026	-18.109	0.556	-5.699	1.410
PFNA	-29.193	0.684	-102.104	3.904	107.124	3.131	-3.436	0.030	-19.339	0.517	-8.271	1.761
PFBS	-20.219	0.582	-72.659	3.592	81.052	2.761	-2.404	0.021	-15.253	0.320	1.024	0.966
PFHxS	-26.044	0.709	-90.264	3.851	97.139	2.958	-3.029	0.033	-17.311	0.603	-4.887	1.457
PFOS	-33.130	1.002	-100.271	3.253	106.197	3.576	-3.556	0.040	-19.828	0.259	-10.934	1.103

Table S9. The hydrogen bond and close contact residues for PFAS ligands bound to different LFABPs. The results were generated by Autodock vina (Trott and Olson 2010) and Autodock tools (Morris et al. 2009).

Ligands	Japanese medaka		Fathead minnow		Human	
	H-bond	Close Contact	H-bond	Close Contact	H-bond	Close Contact
PFBA	ARG121	ILE50, SER52, ARG121	ARG113	ILE42, SER44, ARG113	ARG122, SER124	SER39, ILE62, ILE109, ARG122, SER124
PFPA	ARG121	ILE50, SER52, ASN61, ARG121	ARG113	ILE42, SER44, ARG113	ARG122, SER124	SER39, ILE109, ARG122, SER124
PFHxA	ARG121	ASN61, PHE63, THR73, HIS99, ARG121	ARG113	ILE42, THR65, ARG113	ARG122, SER124	SER39, PHE50, ILE109, ARG122, SER124
PFHpA	ARG121	ILE50, SER52, ASN61, PHE63, THR73, HIS99, ARG121	ARG113	ILE42, SER44, ASN53, PHE55, THR65, HIS91, ARG113	ARG122, SER124	SER39, PHE50, ILE109, ARG122, SER124
PFOA	SER52, ARG121	ILE50, SER52, PHE63, THR73, HIS99, ARG121	ARG113	ILE42, SER44, ASN53, PHE55, ILE63, CYS84, HIS91, ARG113	ARG122, SER124	SER39, PHE50, PHE63, LEU91, THR102, ILE109, ARG122, SER124
PFNA	SER52, ARG121	ILE50, SER52, ASN61, PHE63, ILE71, THR73, CYS92, HIS99, ARG121	ARG113	ILE42, SER44, ASN53, PHE55, ILE63, THR65, CYS84, HIS91, ARG113	ARG122, SER124	SER39, ILE41, PHE50, ILE52, PHE63, LEU91, ARG122, SER124
PFBS	THR73	ILE50, ASN61, PHE63, THR73, HIS99, ARG121	-	ILE42, ASN53, PHE55, HIS91, ARG113	ARG122, SER124	SER39, ASN111, ARG122
PFHxS	ARG121	SER52, ASN61, PHE63, THR73, HIS99, ARG121	ARG113	SER44, ASN53, PHE55, THR65, HIS91, ARG113	ARG122, SER124	SER39, PHE50, ILE109, ARG122, SER124
PFOS	ARG121	ILE50, SER52, ASN61, PHE63, ILE71, THR73, CYS92, HIS99, ARG121	ARG113	ILE42, SER44, ASN53, PHE55, ILE63, THR65, CYS84, HIS91, ARG113	ARG122, SER124	SER39, ILE41, PHE50, LEU91, THR102, ILE109, ARG122, SER124

References

Doering, J. A., Lee, S., Kristiansen, K., Evenseth, L., Barron, M. G., Sylte, I., LaLone, C. A. (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.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. (2009). AutoDock 4 and AutoDockTools 4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30(16), 2785-2791.

Schrodinger, LLC. (2015). The PyMOL molecular graphics system, Version 1.8.

Trott, O., Olson, A. J. (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.