Supplementary Information: Combined *In vitro* and *In silico* Workflow to Deliver Robust, Transparent, and Contextually Rigorous Models of Bioactivity

Nathaniel Charest*, Gabriel Sinclair, Stephanie A. Eytcheson, Daniel Chang, Todd M. Martin, Charles N. Lowe, Katie Paul-Friedman, Antony J. Williams

United States Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, 109 TW Alexander Dr., Research Triangle Park, NC 27711, United States

*Author to whom correspondence should be addressed:
charest.nathaniel@epa.gov

ORCIDs
Nathaniel Charest: 0000-0003-4252-0365
Gabriel Sinclair: 0000-0003-0802-2282
Stephanie Eytcheson: 0000-0002-0007-2421
Daniel Chang: 0000-0001-7561-2747
Todd M. Martin: 0000-0001-5844-8754
Charles Lowe: 0000-0001-9151-6157
Katie Paul-Friedman: 0000-0002-2710-1691
Antony J. Williams: 0000-0002-2668-4821

Supplementary Information 1: Regression Model Investigation

The decision to model an essentially continuous endpoint by discretizing the target and applying a classification algorithm warrants examination. This decision was initially made on intuitive grounds due to the expected degree of empirical variance in the data: translating the data into a categorical model with bins of expected activity was intended to robustly account for latent uncertainty in the data while prioritizing sound interpretation of the model output without complex mathematical constructs. These categorical labels could be translated into activity ranges to approximate the potential bioactivity of molecules when prioritizing chemicals for toxicity testing, while abstaining from overstating the predictive limits of algorithms operating on noisy data.

Main model development focused on the random forest classification model using the binning strategy described in the main text for the above reasons. However, in parallel, we investigated the treatment of the same data with a random forest regression model, following the basic intuition of modeling a continuous endpoint with a continuous algorithm. This investigation generated a notably unexpected result, the mathematical exploration of which may be informative for future QSAR modeling efforts.

Our random forest regressor model was constructed in Python with scikit-learn (https://github.com/mrmsds/ttr-binding-analysis), using the same data under the same splitting as described in the main text for the classifier model. For the regressor, the squared error splitting criterion was applied instead of Shannon information gain, and the same hyperparameter grid was provided, less the class weighting parameter which is not applicable in a regression scenario. The optimal hyperparameters and summary statistics for the parameter-optimized and trained regression model are presented in Tables A2.1 and A2.2.

Hyperparameter	**Options Grid**	**Optimal Value**
n_estimators	50, 100, 500	50
min_samples_leaf	1, 2, 4, 8	4
max_samples	0.33, 0.5, 0.67, 1.0	1.0

Table S1.1: Random forest hyperparameters optimized, with options grid and optimal value.

	Root mean squared error (RMSE) (AU)	**Mean absolute error (MAE) (AU)**	**Median absolute error (MdAE) (AU)**
CV fold 0	24	19	15
CV fold 1	27	21	16
CV fold 2	26	21	17
CV fold 3	24	19	15
CV fold 4	26	21	18
CV SD	1.0	0.89	0.98
CV mean	**25**	**20**	**16**
OOB samples	**26**	**21**	**17**
External test set	**24**	**19**	**16**

Table S1.2: Performance statistics of random forest regression model in internal cross-validation, out-of-bag sample, and external holdout set evaluation.

Recalling that the class bins we defined for the classification model each have a width of 11 AU, in a rough comparison, these statistics appeared similar in performance and stability to those produced by the classifier (cf. Table 3). On discretizing these regressor predictions according to the same scheme and then computing the errors in class units, we obtained a RMSE of 2.2 CU, a MAE of 1.7 CU, and a MdAE of 1.0 CU on the external test set, again very similar to the summary statistics of the classifier.

However, we have proposed in this work that such summary statistics are not meaningfully interpretable for the usability of the model without examination of their context. In this paper, two of our intended contexts were integration with, and eventually replacement of, *in vitro* pre-screening of potentially active compounds to prioritize for concentration-response testing. In the main text, we described a scheme for making such prioritization calls based on the multiclass model predictions, and then compared directly to the *in vitro* experimental prioritization calls made in our coauthors' previous publication, in order to validate the model for this context; therefore, in order to properly evaluate the regression model, we should make the same comparison.

A naïve approach to generate comparable statistics based on regression predictions would be to implement the same 85% threshold used in experimentation in order to binarize the predicted activity values, and then make the same direct comparison to the experimental prioritizations. The results of this approach are displayed in Table A2.3.

	***In silico* prioritized**	***In silico* non-prioritized**	**Summary statistic**
***In vitro* prioritized**	12	31	SN/REC = 0.28
***In vitro* non-prioritized**	2	169	SP = 0.99
Summary statistic	PPV/PRE = 0.86	NPV = 0.85	BA = 0.64

Table S1.3: Confusion matrix and summary statistics of *in silico* vs. *in vitro* compound prioritization outcomes for the regression model with a threshold of 85%.

However, after observing the error statistics of the classification model, we relaxed its prioritization criterion by one class, corresponding to the median absolute error observed across cross-validation and out-of-bag samples. If we observed the same relaxation for the regression model, we would instead implement a threshold value of 69%. Alternatively, we could discretize the predictions and then implement the exact same criterion as we did in classification by prioritizing the predictions which fall in classes 8, 9, and 10. The results of these two approaches are presented in tables A2.4 and A2.5.

	***In silico* prioritized**	***In silico* non-prioritized**	**Summary statistic**
***In vitro* prioritized**	23	20	SN/REC = 0.55
***In vitro* non-prioritized**	9	162	SP = 0.95
Summary statistic	PPV/PRE = 0.72	NPV = 0.89	BA = 0.75

Table S1.4: Confusion matrix and summary statistics of *in silico* vs. *in vitro* compound prioritization outcomes for the regression model with a threshold of 69%.

	***In silico* prioritized**	***In silico* non-prioritized**	**Summary statistic**
***In vitro* prioritized**	18	25	SN/REC = 0.42
***In vitro* non-prioritized**	4	167	SP = 0.98
Summary statistic	PPV/PRE = 0.82	NPV = 0.87	BA = 0.70

Table S1.5: Confusion matrix and summary statistics of *in silico* vs. *in vitro* compound prioritization outcomes for the regression model, with predictions discretized and classes 8, 9, and 10 prioritized.

No matter the criterion we observed, the regression model was notably weaker in context than our multiclass classification model. Not only was it weaker in its overall summary statistic, but more critically, it displayed the exact opposite of our desired error pattern for screening use: it was far more likely to generate false negative than false positive results, meaning this model would be prone to missing potentially active compounds for prioritization. In all but the most lenient of the criteria, in fact, the regression model generated more false negatives than true positives, a clearly unacceptable result.

Continuing our comparison by examining the list of 24 most active compounds identified by experimentation, of which six appeared in our test data set, even under the most lenient criterion (69% threshold), the regression model only prioritized five of them, while the classification model prioritized

all six. Under the most stringent criterion for the regression model (85% threshold), just one of the six was prioritized. Further, examining the numerical activity predictions, all six were predicted more than 20 AU below their measured activities, with two of the six being predicted 30-40 AU below their measured activities; in comparison, the classification model correctly labeled all six compounds as class 10, the highest activity class in our scheme.

On visualizing the distributions of true activity values, classifier predictions, and regressor predictions, utilizing both a continuous kernel density estimate (KDE) plot and discrete histogram according to our predefined bins, a surprising picture emerged (Figure A2.1).

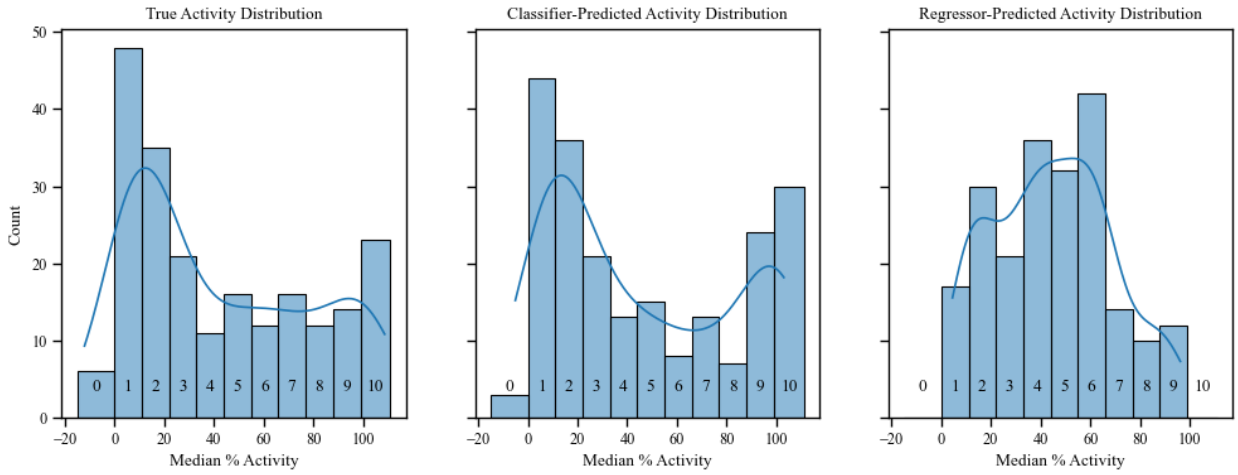


Figure S1.1: Continuous and discrete distribution plots of true activity values, classifier predictions, and regressor predictions.

As noted in the main text, the empirical data we modeled here exhibited a marked bimodal distribution. The main multiclass random forest classification model captured this distribution with relative fidelity; where it made errors, those errors tended to be in the form of over-emphasizing the bimode and over-predicting active compounds (complete discussion of these error patterns in main text). On the other hand, the random forest regression model, operating on exactly the same data, with the closest possible optimization and training process for the context, and generating nearly identical score statistics, abjectly failed to capture the same underlying distribution.

We attribute this regression-specific error pattern chiefly to the aggregation function implemented by random forest regressors (or indeed most ensemble and distance-based regressors). In a random forest regression model, the per-tree predictions are generated by taking the arithmetic mean of training set values in each terminal leaf, and the final prediction is generated by taking another arithmetic mean of these predictions. On the other hand, in a random forest classifier model, the predictions are generated at each stage by majority (or plurality) voting on the class outcome. The use of the arithmetic mean reduces the robustness of the algorithm to outliers (in this case, structural misclassifications by individual nodes or trees) and embeds a fundamental assumption of a unimodal distribution which does not hold for this data set.

To validate the hypothesis that this pattern was truly a mathematical artefact of the algorithm and not a consequence of flawed regression model training or some other experimental error, we considered the toy example of a *k*-nearest neighbors (kNN) classifier and regressor. A kNN is a less complex and more transparent algorithm than a random forest, and we used its transparency to demonstrate in clear terms

how the mathematical aggregation function alone led to the same error pattern seen in the random forest model.

Because the selection of neighbors by a kNN follows deterministically from the distance calculations on the training and test set, a kNN classifier and kNN regressor trained on the same data, parameterized by the same distance metrics, must produce the exact same selection of training set neighbors for each test set query point. In order to demonstrate this, we trained a kNN classifier and regressor, using $k = 5$ and a brute-force algorithm based on Euclidean distances for both learners, using Python scikit-learn (https://github.com/mrmsds/ttr-binding-analysis). We then compared the *k*-neighbors graphs generated by these learners in order to verify that the same neighbors were in fact selected for each point.

Once the training set neighbors of each point are selected, the kNN algorithm calculates a final prediction by taking a majority (or plurality) class vote among the neighbors (in the case of a classifier) or the arithmetic mean of the target variable among the neighbors (in the case of a regressor), precisely analogous to the aggregation used within and between trees in a random forest. Because we verified that the selected neighbors were the same for both learners, any differences in the resulting distributions were solely attributable to differing properties of the mathematical functions used for aggregation; there was no possibility of insufficient training or poor parameterization leading to different outcomes.

Just as in the comparison of the random forest classifier and regressor outcomes, we plotted the distributions of true activity values, kNN classifier predictions, and kNN regressor predictions, utilizing both a continuous kernel density estimate (KDE) plot and discrete histogram according to our predefined bins (Figure A2.2).

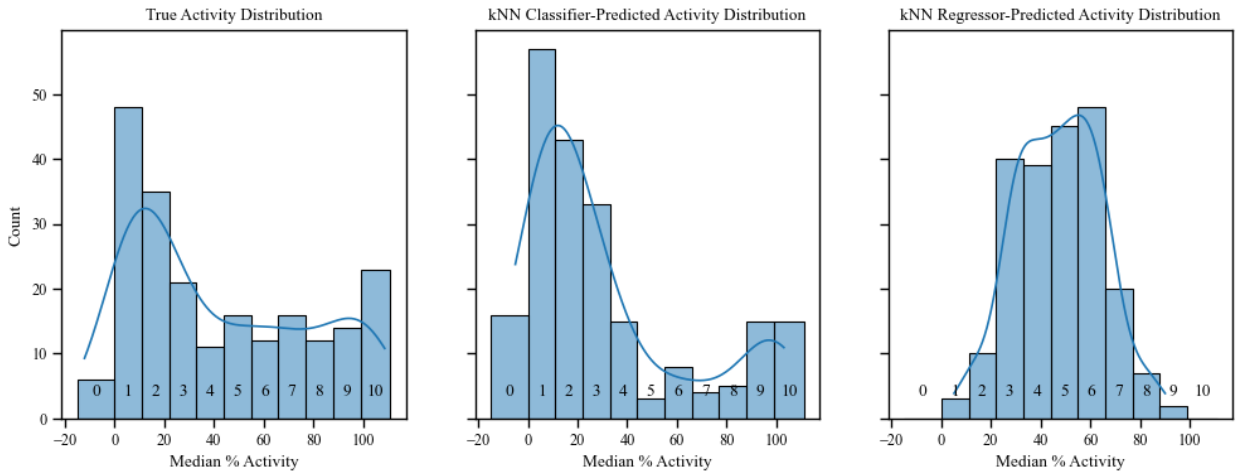


Figure S1.2: Continuous and discrete distribution plots of true activity values, kNN classifier predictions, and kNN regressor predictions.

Despite controlling for all other differences in the models, the kNN classifier and regressor exhibited the same pattern seen in the random forest classifier and regressor. The classifier reproduced the bimodal distribution of the underlying data with relative fidelity, while the regressor forced it into a unimodal distribution which did not reflect reality.

Summary statistics for both the kNN classifier and kNN regressor on the external test data are presented in Table A2.6; the performance of the regressor appeared to be improved over the classifier, despite its failure to capture the true distribution of the data. Just as in the case of the random forest regressor, this

would be missed by examining rote score statistics without a complete understanding and scrutiny of the true data and prediction distributions.

	kNN classifier	**kNN regressor**
Root mean squared error (RMSE) (CU)	4.2	3.2
Mean absolute error (MAE) (CU)	3.1	2.6
Median absolute error (MdAE) (CU)	2.0	2.0

Table S1.6: Summary statistics of the kNN classifier and kNN regressor on the external test data set in class units (CU).

An exhaustive investigation of this phenomenon in bioactivity and physicochemical property QSARs is beyond the scope of this paper, and the explanation articulated here represents a hypothesis for further investigation, not a unilateral claim on this data set or any other. However, these computational experiments support the thesis we articulate in this work, that the definition of an explicit purpose context, and the evaluation of a model with experiments and statistics targeted to that purpose context, is critical to truly validate the model for usage; and we further emphasize the value of thorough understanding of the mathematical foundation and properties of machine learning algorithms and the visualization of feature and endpoint distributions.

Supplementary Information 2: Docking simulation and analysis of the human TTR binding pocket domain for 2ROX and 3CFN

This supplemental information section details the docking simulations and protein-ligand interaction analysis performed to support the feature engineering aspect of the main study.

Computational tools

Docking studies and protein-ligand interaction analyses were performed using the Molecular Operating Environment software platform, MOE 2024.06 (www.chemcomp.com and [1]).

Active-site definition and protein preparation

The three-dimensional *holo* or ligand-bound structures of human transthyretin with the 8-anilino-1-naphthalene sulfonate (ANSA) [3CFN, 1.87 Å resolution] and the endogenous thyroxine (T4) [2ROX, 2.00 Å resolution] were retrieved from the RCSB protein data bank (www.rcsb.org). A biomolecular transformation was applied to generate the homotetramer bioassembly (i.e., chains A, B, C and D) which resulted in 2 distinct "active sites" formed at the interface of each homodimer unit. For each ligand bound structure, the two "active sites" were defined by each residues' proximal distance to the ligand within 4.5 A. Figure S1 and S2 depict each active site relative to the crystallographic bound ligand. For consistency of the designation, the active site "1" is composed of receptor chains A and C and the active site "2" is composed of receptor chains B and D. Details on the receptor chain sequences are provided in the Supplemental Data file.

For simulation purposes (i.e., molecular docking), receptor atoms (protein backbone and sidechains) were fixed beyond 8 A from the ligand and tethered (i.e., constrained) with a force constant to ensure that receptor atoms do not deviate too much from their initial crystallographic positions esp. during any geometry optimization step. Structure preparation is necessary to ensure stable simulation of the system and encompasses several steps including protonation of all relevant residues, proper assignment of charges, treatment of bound water, correction of missing residue atoms, and capping termini as well as performing a constrained geometry optimization on the receptor as defined by fixed and tethered atom constraints. A hybrid Amber:EHT atomic forcefield was utilized to parameterize both protein (using the AMBER 19 forcefield [2]) and non-standard ligand atom types (via extended Huckel Theory [3]).

Molecular docking

Classical forcefield docking simulations were performed in 2 steps via MOE with the applied "active site" definition for both ANSA and T4. First, a set of ligand poses were generated through initial placement and subsequent scoring based on the London dG scoring function which estimates the binding free energy of a ligand pose as a summation of Coulombic interactions, solvation effects, van der Waals interactions and surface area contributions. In the second step, a constrained re-optimization of flexible residue sidechains (tethered) near the ligand (i.e., an "induced fit" approximation) was performed which is then re-scored using a more accurate parameterized Generalized-Born Volume Integral/Weighted Surface Area scoring function which penalizes exposed surface areas [4,5] (GBVI/WSA dG). A maximum of 500 initial placements were retained

in the first step and a maximum of 200 induced fit poses were refined and scored for the final docked poses in each active site.

Protein-ligand Interactions

Protein-Ligand Interaction Fingerprints (PLIF) summarize the interactions between protein and ligands using a bit-wise representation for specific interaction types between a protein residue and the ligand. The types of interactions are typically characterized by hydrogen bonding (i.e., acceptor or donor to either the residue sidechain or protein backbone), ionic interactions, arene interactions and solvent interactions. For this study, PLIFs were generated for each ligand docked pose within each active site (*i.e.*, 1 or 2). For the 2 bound ligands, 400 and 333 induced fit docked poses were generated for T4 and ANSA, respectively. For each docked pose with their receptor target, PLIFs are generated and stored as a raw data (i.e., interaction energies in kcal/mol) score for the interaction type. In this formalism, a "bit" is a bounded interpretation of the minimum energy threshold – i.e., weak interactions (i.e., Arene1) are defined by a minimum energy threshold and strong interactions (i.e., Arene2) are defined by a higher minimum energy threshold/cutoff. By default, these thresholds are generated after final docking poses are scored using the default thresholds of weak and strong interactions given in Table S1. The raw data (i.e., interaction energies in kcal/mol) can be used with the aforementioned cutoffs to generate bit-wise representations (PLIFs) and are given in the Supplemental Data file.

Table S2.1 Default thresholds (kcal/mol) for creating bit-wise representations (i.e., PLIFs) used in MOE.

Interaction Type	**(1) Weak cutoffs**	**(2) Strong cutoffs**
Sidechain H-donor	0.5	1.5
Sidechain H-acceptor	0.5	1.5
Backbone H-donor	0.5	1.5
Backbone H-acceptor	0.5	1.5
Solvent H-donor	0.5	1.5
Solvent H-acceptor	0.5	1.5
Ionic Attraction	0.5	3.5
Arene Attraction	0.5	1.0

Analysis & discussion

For each active site/ligand pair, the docked poses were inspected and filtered for specific interactions based on the observed interaction types and energy threshold (i.e., weak/strong). For pragmatic reasons, solvent interactions were not included in the analysis since the goal was to identify critical non-water residue interactions that would assist in elucidating the binding motifs observe between the protein and the studied ligands (ANSA and T4).

For ANSA docked results, 20 interaction types across 8 residue positions in both active sites were identified. Not surprisingly, both active sites exhibit similar interaction fingerprints primarily with LYS15 on both chains. Ionic, sidechain acceptor and arene-type interactions were observed to be the most abundant in the LYS15 interactions with a maximum observed abundance of ~67–68%

for ionic-type interactions, ~58–59% for sidechain acceptor type interactions and ~19–33% for arene-type interactions in all docked poses. While LYS15 is particularly interesting as it resides on the interior beta sheet at the interface of the 2 monomeric units, both LEU17 and VAL121 play a smaller part in stabilizing the ligand through Arene-type interactions.

For the T4 docked poses, a more complex interaction than ANSA was observed with 58 interactions across 19 non-water residues and 53 interactions across 18 non-water residues identified in active sites 1 and 2, respectively. This difference in the number and types of interactions as well as residues between active sites may be attributed to previously observed negative-cooperativity [6] effects in TTR initiated by the endogenous ligand interaction with the SER117 residue. For the purposes of this study, this effect was not investigated further. Similar to the ANSA ligand, T4 interactions with LYS15 on each chain of the active site pocket were observed to be the most abundant in the docked poses. Unlike ANSA, arene-type interactions were observed in a higher percentage of docked poses – i.e., 46% of interactions as either being attributed to LYS15 on either chain A or C in active site 1 and 54% of interactions attributed to LYS15 on either chain B or D in active site 2. Additional residue interactions can be attributed to an Ionic interaction with LYS15 (21%) and a sidechain acceptor interaction with LYS15 (14.5%). Residue, interaction type, and relative abundance in the docked poses are summarized for each site and ligand in the Supplemental Data file, including for completeness water co-crystallized water interactions already present in each protein-ligand complex (2ROX and 3CFN). A summary of the top 6 interactions with abundance, type and chain residue ID are displayed in Tables S2 and S3 for each active site of T4 and ANSA, respectively.

Table S2.2 Top 6 interactions based on % abundance in the docked poses for each TTR:T4 complex and active site excluding water-ligand interactions.

Chain Residue ID | Interaction Type	2ROX T4 Site 1	Chain Residue ID | Interaction Type	2ROX T4 Site 2
C15 | Arene1	23.5%	B15 | Arene1	27.0%
A15 | Arene1	22.5%	D15 | Arene1	27.0%
A13 | ChDon1	15.0%	B15 | Ionic1	21.0%
A104 | Ionic1	15.0%	B15 | ChAcc1	14.5%
C13 | ChDon1	14.5%	D54 | Ionic1	14.0%
A104 | ChAcc1	14.5%	D54 | ChDon1	14.0%

Table S2.3 Top 6 interactions based on % abundance in the docked poses for each TTR:ANSA complex and active site excluding water-ligand interactions.

Chain Residue ID | Interaction Type	2ROX ANSA Site 1	Chain Residue ID | Interaction Type	2ROX ANSA Site 2
A15 | Ionic1	67.0%	D15 | Ionic1	68.8%
A15 | ChAcc1	59.2%	D15 | ChAcc1	58.4%

A15 | Ionic2	54.7%	D15 | Ionic2	52.6%
A15 | ChAcc2	50.8%	D15 | ChAcc2	51.9%
C15 | Ionic1	39.1%	B15 | Ionic1	46.1%
A15 | Arene1	32.9%	B15 | ChAcc1	26.0%
C15|ChAcc1	25.1%	D15 | Arene1	19.5%

Based on the above analysis for abundance and inspection of docked poses, the LYS15 residue was identified as being significant in describing the binding domain of both T4 and ANSA. However, distinct differences in the binding patterns (i.e., interaction types) are observed through the PLIFs as we see 2 distinct binding modes for T4 and ASNA. Based on docked poses, PLIFs and the observed crystal structures (3CFN and 2ROX) of the two bound ligands, we can deduce that arene type interactions may be important in stabilizing the endogenous ligand (T4) in the binding pocket. While the ANSA binding mode may depend more heavily on a combination of Ionic and H-bonding interactions to stabilize itself in the pocket, arene-type interactions still exist with both LYS15 and LEU17 playing a smaller role in stabilizing ANSA. For competitive binding, ANSA vs. T4, additional ligand stability may be gained from interactions that stabilize the ring structures of T4 relative to ANSA as are observed in the LYS15 arene-type interactions.

References

1. *Molecular Operating Environment (MOE)*, 2024.06; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2024.
2. Tian, C., Kasavajhala, K., Belfon, K., Raguette, L., Huang, H., Migues, A., Bickel, J., Wang, Y., Pincay, J., Wu, Q., Simmerling, C.; ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J. Chem. Theory Comput. (2020) 528–552.
3. Gerber, P.R., Müller, K.; MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry; J. Comput. Aided Mol. Des. Jun 9(3) (1995) 251–68.
4. Corbeil, C.R., Williams, C.I. & Labute, P. Variability in docking success rates due to dataset preparation. J Comput Aided Mol Des 26, 775–786 (2012). https://doi.org/10.1007/s10822-012-9570-1
5. Labute, P. The Generalized Born / Volume Integral (GB/VI) Implicit Solvent Model: Estimation of the Free Energy of Hydration Using London Dispersion Instead of Atomic Surface Area; J. Comp. Chem. 19 (2008) 1693–1698.
6. Tomar D, Khan T, Singh RR, Mishra S, Gupta S, et al. (2012) Crystallographic Study of Novel Transthyretin Ligands Exhibiting Negative-Cooperativity between Two Thyroxine Binding Sites. PLOS ONE 7(9): e43522. https://doi.org/10.1371/journal.pone.0043522

SI2.1 Figure & Figure captions

Figure S1.

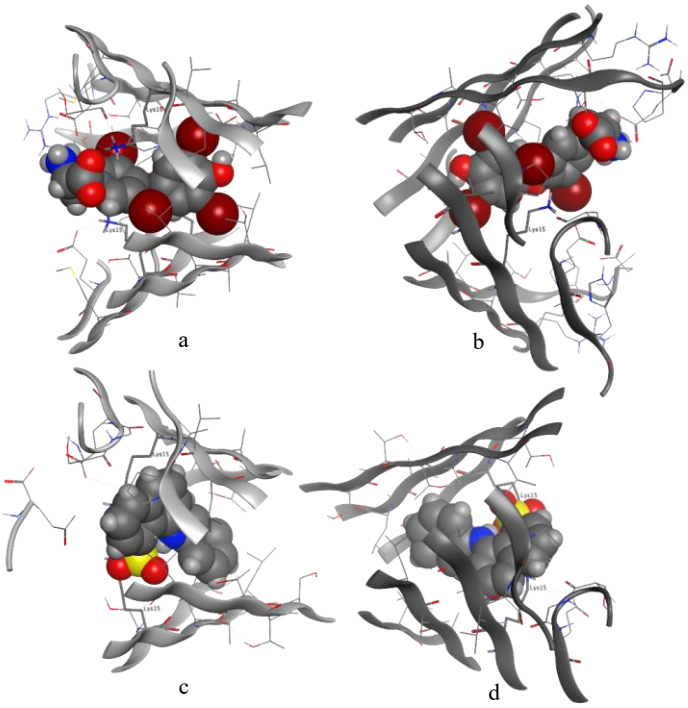


Figure S1. A 3D depiction of active site 1 and 2 (the ligand is represented as a space filling model, select residues are represented as stick models and the protein backbone is depicted with a ribbon representation) for: a) T4 in active site 1, b) T4 in active site 2, c) ANSA in active site 1, and d) ANSA in active site 2. TTR receptor chains 1 and 3 are light grey and depicted in a and c, while receptor chains 2 and 4 are dark grey and depicted in b and d to illustrate the construction of the active site at the interface of each respective chain combination.

Figure S2.

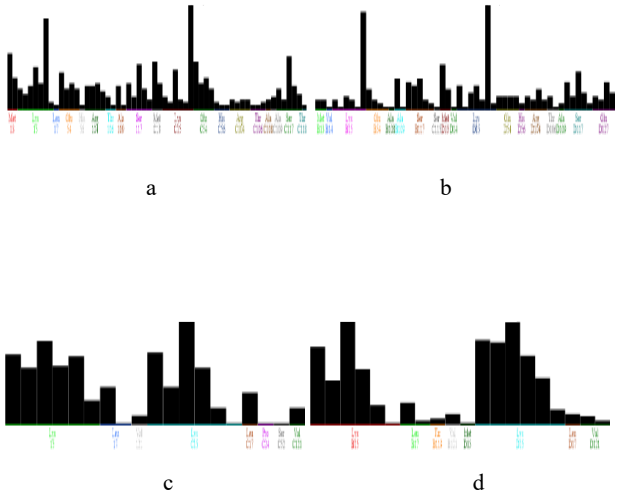


Figure S2.2 Normalized population statistics (as a histogram) are provided for key TTR residue interactions with the ligand for the series of docked poses within each active site. Labels for each residue are provided on the x-axis of each plot. Each histogram bar represents a specific interaction (i.e., Acceptor, Donor, Ionic, Arene, etc) for each identified residue. The population statistics are represented for docked poses for: a) T4 in active site 1, b) T4 in active site 2, c) ANSA in active site 1, and d) ANSA in active site 2.

Supplementary Information 3: Micelle and Autofluorescence Experiments

Fifty-three chemicals from ToxCast chemical libraries (ph1_v2, ph2, and e1k) were screened to determine whether formation of micelles may result in false positives in a fluorescence-based TTR binding assay (1). The chemicals were selected by running the list of ToxCast chemicals through a structure-based classification model to identify likely surfactants and chemicals with zero predicted surfactant likelihood (2). In addition, X-ray crystal structures existed for several of the chemicals bound to TTR, which identified them as likely "true positives". In total, the set of 53 chemicals screened included 16 predicted likely surfactants, 2 PFAS, the known surfactant sodium dodecyl sulfate, 31 predicted non-surfactants, and 3 "true positives" (Table S3.1).

Chemicals were screened in three different scenarios: one assessing chemical autofluorescence ("autofluorescence"), one to assess whether ANSA fluoresces due to micelle formation ("micelle"), and one screening for TTR binding ("TTR"). Screening in these additional assays closely followed methods described in Eytcheson et al. (2024) (3). Three replicate reaction plates were set up for each scenario. Master mixes were prepared as outlined in Table A3.2. Each reaction plate contained a T4 standard curve and several control wells (shaded gray in the table), including background fluorescence of TTR and ANSA. Chemicals were provided by Evotec (Branford, Connecticut) in 96-well plates at a target concentration of 20 mM in DMSO. Reaction plates were loaded with the appropriate amount of master mix. Subsequently, T4 and test chemical were loaded to each reaction plate using a Liquidator 96-channel benchtop pipettor (Mettler Toledo, Columbus, Ohio). Immediately following addition of T4 and test chemicals, the reaction plates were sealed with a polyolefin plate seal and shaken on a Jitterbug microplate shaker (Boekel Scientific, Feasterville, Pennsylvania) at 1000 rpm for 2 minutes. Plates were covered and incubated in the dark at 4 °C for 2 hours. Immediately after incubation, fluorescence was measured in each plate using a BioTek Synergy Neo2 plate reader (Agilent, Santa Clara, California) with an excitation of 380 nm and emission of 475 nm.

DTXSID	CASRN	Preferred Name	Notes
DTXSID0041567	141-22-0	(9Z,12R)-12-Hydroxyoctadec-9-enoic acid	Predicted as very likely surfactant (>0.95)
DTXSID4042416	1191-50-0	Sodium myristyl sulfate	Predicted as very likely surfactant (>0.95)
DTXSID0042169	5116-94-9	Monotridecyl phosphate	Predicted as very likely surfactant (>0.95)
DTXSID0042400	1120-01-0	Sodium hexyldecyl sulfate	Predicted as very likely surfactant (>0.95)
DTXSID6047103	1120-04-3	Octadecyl sulfate sodium salt	Predicted as very likely surfactant (>0.95)
DTXSID5037028	57-09-0	Hexadecyltrimethylammonium bromide	Predicted as very likely surfactant (>0.95)
DTXSID9021554	334-48-5	Decanoic acid	Predicted as very likely surfactant (>0.95)
DTXSID1025809	112-80-1	Oleic acid	Predicted as very likely surfactant (>0.95)
DTXSID7025506	463-40-1	Linolenic acid	Predicted as very likely surfactant (>0.95)
DTXSID5042243	110-25-8	Oleyl sarcosine	Predicted as very likely surfactant (>0.95)
DTXSID2025505	60-33-3	Linoleic acid	Predicted as very likely surfactant (>0.95)
DTXSID8021642	57-11-4	Octadecanoic acid	Predicted as very likely surfactant (>0.95)
DTXSID5021590	143-07-7	Dodecanoic acid	Predicted as very likely surfactant (>0.95)
DTXSID7042011	97-78-9	N-Dodecanoyl-N-methylglycine	Predicted as very likely surfactant (>0.95)
DTXSID2021602	57-10-3	Hexadecanoic acid	Predicted as very likely surfactant (>0.95)
DTXSID6021666	544-63-8	Tetradecanoic acid	Predicted as very likely surfactant (>0.95)
DTXSID3031864	1763-23-1	Perfluorooctanesulfonic acid	Check if surfactant-like PF_S trigger fluorescence
DTXSID3031862	307-24-4	Perfluorohexanoic acid	Check PFHA
DTXSID1026031	151-21-3	Sodium dodecyl sulfate	Surfactant originally reported to cause ANSA fluorescence
DTXSID1020516	57-14-7	1,1-Dimethylhydrazine	Predicted 0 Surfactant
DTXSID7027461	2231-57-4	Thiocarbazide	Predicted 0 Surfactant
DTXSID7044390	6610-29-3	N-Methylhydrazinecarbothioamide	Predicted 0 Surfactant
DTXSID0021464	137-30-4	Ziram	Predicted 0 Surfactant
DTXSID1029677	7631-86-9	Silica	Predicted 0 Surfactant
DTXSID2041125	7681-82-5	Sodium iodide	Predicted 0 Surfactant
DTXSID4029692	7757-79-1	Potassium nitrate	Predicted 0 Surfactant

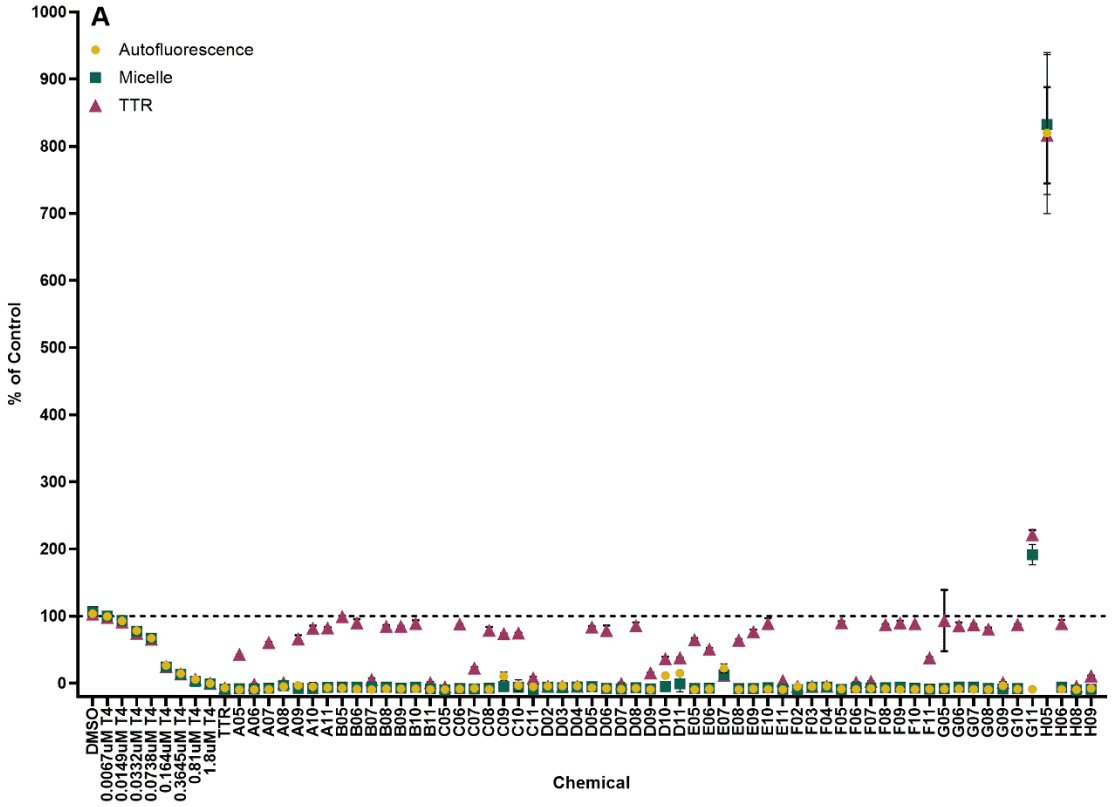
DTXSID5020811	7487-94-7	Mercuric chloride	Predicted 0 Surfactant
DTXSID5023825	1303-11-3	Indium arsenide	Predicted 0 Surfactant
DTXSID8020121	26628-22-8	Sodium azide	Predicted 0 Surfactant
DTXSID8041909	10025-74-8	Dysprosium (III) chloride	Predicted 0 Surfactant
DTXSID9021348	62-56-6	Thiourea	Predicted 0 Surfactant
DTXSID1020354	461-58-5	Cyanoguanidine	Predicted 0 Surfactant
DTXSID2029167	137-42-8	Metam-sodium	Predicted 0 Surfactant
DTXSID4026769	107-46-0	Hexamethyldisiloxane	Predicted 0 Surfactant
DTXSID7020508	75-60-5	Dimethylarsinic acid	Predicted 0 Surfactant
DTXSID0042379	1112-39-6	Dimethoxydimethylsilane	Predicted 0 Surfactant
DTXSID9040710	107-51-7	Octamethyltrisiloxane	Predicted 0 Surfactant
DTXSID1038795	1873-88-7	1,1,1,3,5,5,5-Heptamethyltrisiloxane	Predicted 0 Surfactant
DTXSID2025395	999-97-3	Hexamethyldisilazane	Predicted 0 Surfactant
DTXSID2040361	6734-80-1	Metam-sodium hydrate	Predicted 0 Surfactant
DTXSID2027204	556-61-6	Methyl isothiocyanate	Predicted 0 Surfactant
DTXSID6027050	128-04-1	Sodium dimethyldithiocarbamate	Predicted 0 Surfactant
DTXSID6027131	288-88-0	1H-1,2,4-Triazole	Predicted 0 Surfactant
DTXSID1020194	10043-35-3	Boric acid (H3BO3)	Predicted 0 Surfactant
DTXSID7021029	62-75-9	N-Nitrosodimethylamine	Predicted 0 Surfactant
DTXSID2042191	534-13-4	N,N'-Dimethylthiourea	Predicted 0 Surfactant
DTXSID9033058	584-13-4	4-Amino-1,2,4-triazole	Predicted 0 Surfactant
DTXSID8025599	6317-18-6	Methylene bis(thiocyanate)	Predicted 0 Surfactant
DTXSID5021332	137-26-8	Bis(dimethylaminothiocarbonyl) disulfide	Predicted 0 Surfactant
DTXSID7027205	556-67-2	Octamethylcyclotetrasiloxane	Predicted 0 Surfactant
DTXSID4021218	117-39-5	Quercetin	Crystal Structure
DTXSID5022308	446-72-0	Genistein	Crystal Structure
DTXSID5041306	131-55-5	2,2',4,4'-Tetrahydroxybenzophenone	Crystal Structure

Table S3.1: Chemicals screened in supplementary experimentation and reasons for selection.

Reagent	Concentration	T4	TTR	ANSA	Autofluorescence	Micelle	TTR
TTR	0.125 μM	X	X				X
ANSA	1.2 μM	X		X		X	X
DMSO	0.5%	X	X				*
Phosphate buffer	NA	X	X	X	X	X	X
T4	Varies	X					
Test chemical	Target 100 μM				X	X	X

Table S3.2: Reaction components. (Chemicals are in DMSO, so no additional DMSO was added to the master mix.)***

Results from screening in these additional assays are plotted in Figure A1.1 as % of control as calculated relative to the high (1.8 μM) and low (0.0067 μM) T4 concentrations. In the autofluorescence and micelle assays, 0% of control indicates no fluorescence measured in the well (*i.e.*, no chemical autofluorescence or ANSA fluorescing within a micelle). For TTR-binding, activity near 0% of control indicates the chemical has bound to TTR and displaced ANSA. Results from screening for autofluorescence indicated that the chemical in well H05 (2,2',4,4'-tetrahydroxybenzophenone, selected for screening based on crystal structure) was highly autofluorescent (~8.5X fluorescence in the DMSO negative control; Figure S1A). With well H5 excluded, other chemicals exhibiting some autofluorescence included C09, D10, D11, and E07 (Figure S1B). Two chemicals exhibited high fluorescence in the screen for micelles: G11 (hexadecyltrimethylammonium bromide, a predicted likely surfactant) and H05 (2,2',4,4'-tetrahydroxybenzophenone, identified as autofluorescent). Overall, screening in the TTR binding assay replicated results from the previous screening effort (Eytcheson, et al. 2024), and while autofluorescence may confound interpretation of TTR binding for a few chemicals, micelle formation does not seem to occur for a majority of chemicals at the concentration used in the assay.


A
Autofluorescence
Micelle
TTR
% of Control
Chemical
DMSO
0.0067uM T4
0.0149uM T4
0.0332uM T4
0.0738uM T4
0.164uM T4
0.3645uM T4
0.81uM T4
1.8uM T4
TTR

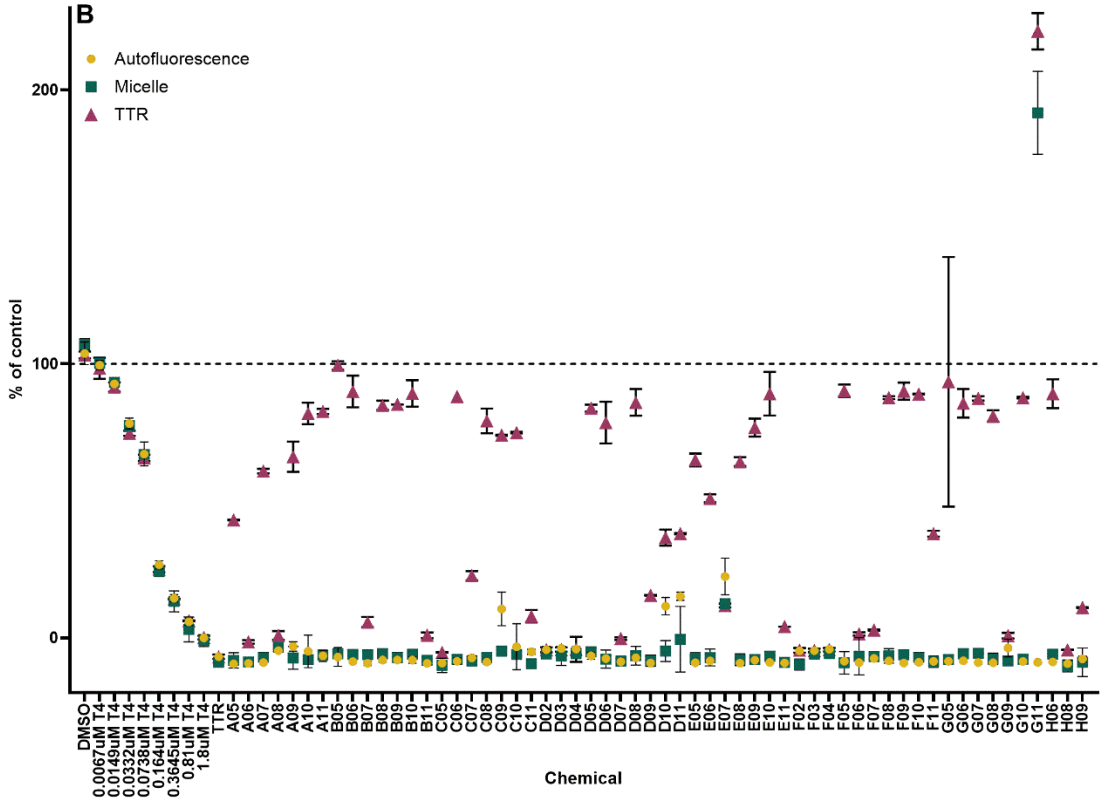

B
Autofluorescence
Micelle
TTR
% of control
Chemical
DMSO
0.0067uM T4
0.0149uM T4
0.0332uM T4
0.0738uM T4
0.164uM T4
0.3645uM T4
0.81uM T4
1.8uM T4
TTR

Figure S3.1: Activity of 53 chemicals screened in three different scenarios: autofluorescence (yellow circle), micelle formation (green square), and TTR binding (pink triangle). Results include DMSO negative control and a T4 curve with the lowest concentration of T4 (0.0067 µM) representing maximum assay fluorescence (100% control) and the highest concentration of T4 (1.8 µM) representing complete displacement of ANSA from TTR (0% of control). Figure 1A depicts all data, and Figure 1B excludes data from well H05 (2,2',4,4'-Tetrahydroxybenzophenone).

References

(1) De Vendittis, E., Palumbo, G., Parlato, G., & Bocchini, V. (1981). A fluorimetric method for the estimation of the critical micelle concentration of surfactants. *Anal Biochem, 115*(2), 278-286. doi:10.1016/0003-2697(81)90006-3

(2) Phillips, K. A., Wambaugh, J. F., Grulke, C. M., Dionisio, K. L., & Isaacs, K. K. (2017). High-throughput screening of chemicals as functional substitutes using structure-based classification models. *Green Chem, 19*(4), 1063-1074. doi:10.1039/C6GC02744J

(3) Eytcheson, S. A., Zosel, A. D., Olker, J. H., Hornung, M. W., & Degitz, S. J. (2024). Screening the ToxCast Chemical Libraries for Binding to Transthyretin. *Chem Res Toxicol*. doi:10.1021/acs.chemrestox.4c00215

Supplementary Information 4: Activity of Per- and Polyfluoroalkyl Substances

An interesting trend noted in the TTR data from Eytcheson et al. was the relatively high activity of per- and polyfluoroalkyl substances (PFAS). PFAS are especially relevant in current regulatory activities and scientific research, and are of concern to the underlying toxicological interest of this work: TTR transport is hypothesized to be more significant in human fetal development than in adult humans due to its greater prevalence in fetal matrices, and as certain PFAS are capable of crossing the placental barrier in humans (1), previous investigations have drawn attention to the potential for developmental impacts of this mechanism of TH interference (2).

Of the chemicals tested in the Eytcheson data, 19 were listed in the EPA's list PFAS8a7v3, available on the CompTox Dashboard (3), consisting of the subset of reportable substances that meet the TSCA section 8(a)(7) definition of PFAS chemicals. The QC process outlined in the prior section rejected 15 of these chemicals from the training data set (two due to explicit failure calls and the remainder due to unknown quality calls). The two failures were due to unexpectedly low purity or concentration in the samples after incubation, indicating potential degradation of the analytes. For the remaining unknowns, the QC data could not identify presence or purity based on the applied analytical methods, and thus the quality of the samples was characterized as unknown. The four PFAS compounds which passed QC, in the process of data set splitting, were all randomly assigned to the internal training set.

Despite omitting the majority of PFAS compounds from QSAR model training and scoring due to their inadequate analytical QC data, we deemed it important to investigate them, albeit tentatively, due to their status as key emerging contaminants, and due to the existing literature modeling interaction of these compounds with TTR using QSAR modeling, molecular docking, and/or molecular dynamics simulation (2,4,5). The RCSB protein database contained crystallographic structures with bound perfluorooctanoic acid (PDB: 5JID, Figure 6) (6), shown with its interaction map. This crystal structure exhibits a binding motif broadly consistent with that of the endogenous ligand T4: contact by external residue LYS15 with the polar headgroup and insertion of the non-polar surface into the beta sheet cleft. An identical binding motif is observed for perfluorooctanesulfonic acid (PDB: 5JIM) (6). Electronegative repulsion can be expected to drive the extended conformer of perfluoroalkane chains to dominate, the surface of which shows favorable interactions with the beta sheet cleft (5). The observation that the anionic headgroup tends to face externally is consistent with prior *in silico* exploration of PFAS systems binding to TTR (5).

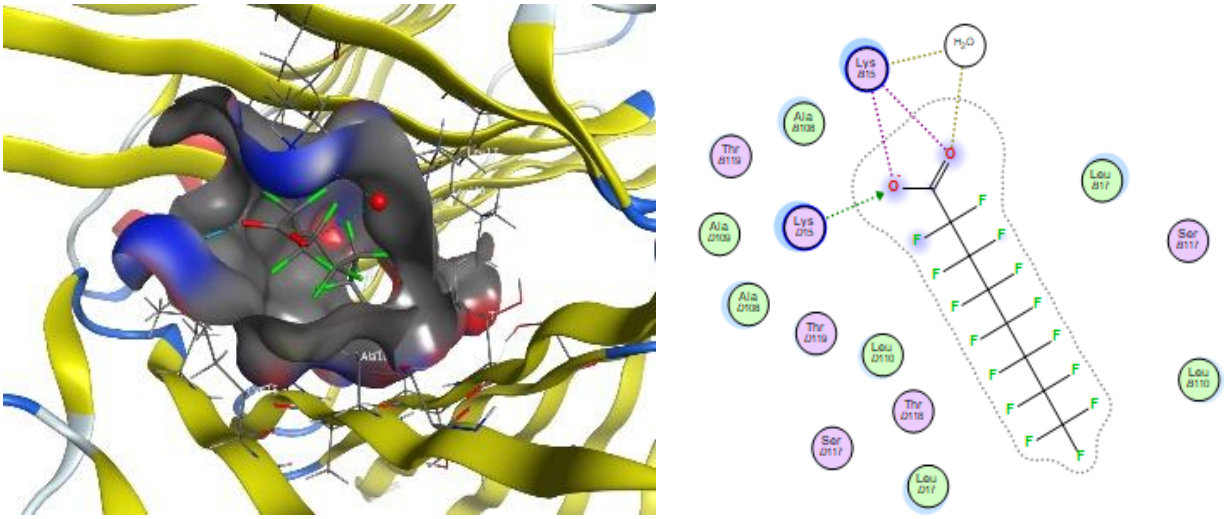


Figure S4.1: The binding of linear perfluorooctanoic acid to the binding site of TTR. The electrostatic repulsion of fluorine minimizes the conformational entropy of the perfluoroalkyl chain and creates a stable hydrophobic interaction with the beta sheets, while the acidic headgroup contacts LYS15.

The predictions of our model on the PFAS compounds with "unknown" QC rating in the data set are presented in Table S4.1. (The compounds which outright failed QC have been omitted.) For comparison, we also present the measurements of Weiss et al. (2009), who directly measured the inhibitory activity of perfluorinated compounds on T4 at 10 µM. These measurements are not directly comparable to those of Eytcheson et al., as the fluorescence assay relies on displacement of ANSA, a weaker TTR binder than T4; however, they provide independent confirmation of the high activity of several of these compounds, which was both seen in experiments by Eytcheson et al. and predicted by our model.

DTXSID	Chemical name	Experimental median activity (%)	T4 inhibition at 10 uM (Weiss 2009) (%)	Experimental class	Predicted class
DTXSID8031865	Perfluorooctanoic acid	103.5	96	10	10
DTXSID3031864	Perfluorooctanesulfonic acid	104.9	99	10	10
DTXSID8037706	Potassium perfluorooctanesulfonate	104.3	Not tested	10	10
DTXSID3031860	Perfluorodecanoic acid	102.9	54	10	10
DTXSID7029904	Fluorotelomer alcohol 8:2	45.8	-17	5	10
DTXSID8031863	Perfluorononanoic acid	104.7	82	10	10
DTXSID3037709	Potassium perfluorohexanesulfonate	102.4	Not tested	10	10
DTXSID3037707	Potassium perfluorobutanesulfonate	99.4	Not tested	10	10
DTXSID1037303	Perfluoroheptanoic acid	102.2	93	10	10
DTXSID8047553	Perfluoroundecanoic acid	101.4	26	10	10
DTXSID3031862	Perfluorohexanoic acid	96.4	57	9	10
DTXSID3047558	2-(Perfluorohexyl)ethyl methacrylate	9.4	Not tested	1	3
DTXSID8037708	Ammonium perfluorooctanoate	104.2	Not tested	10	10

Table S4.1: Experimental and predicted activity values of PFAS compounds for TTR binding and T4-TTR binding inhibition.

Our model achieved predictions within one class of true for all but two of 13 compounds, and successfully prioritized all compounds that were prioritized for concentration-response testing by the *in vitro* assay. We also noted two pairs of compounds which collapsed to the same QSAR-ready structure on removal of counterions (perfluorooctanoic acid/ammonium perfluorooctanoate and perfluorooctanesulfonic acid/potassium perfluorooctanesulfonate), which were identically predicted by the model and experimentally shown to have substantially the same activity *in vitro*. This affirmed that the method by which QSAR-ready structures were derived was appropriate for this model.

The first of two compounds with >1 class error, fluorotelomer alcohol 8:2, was predicted as most active (class 10), but had a moderate experimental median activity of 45.2% This suggests that our model may have classified the hydroxyl group as a hydrophilic head group analogous to the carboxylic and sulfonic acids which give the majority of other tested PFAS compounds their activity, when in reality this group cannot form hydrogen bonds as effectively with external residue LYS15 or GLU54, reducing its stability in the TTR binding pocket and preventing complete inhibition. Future iterations of the model could utilize more sampling of the fluorotelomer alcohols and/or testing of other descriptor sets, such as circular fingerprints, to better resolve functional group differences and reduce the likelihood of this form of error.

The second such compound, 2-(perfluorohexyl)ethyl methacrylate, was identified by our model as having low moderate activity (class 3), when it was experimentally inactive (class 1), likely due to the insulation of its hydrophilic head group by the methacrylate ester preventing it from appreciably contacting the relevant external residues. It is reassuring that our model correctly identified this compound as having uniquely low activity among the PFAS species tested, although the prediction was not precise.

We emphasize once again that these 13 compounds were not reliably quantified under our analytical QC process, although they did not present critical problems as did the two that failed the process entirely. Therefore, these compounds were not included in the internal data set on which the model was trained, nor the external set on which its performance was evaluated. These results are presented separately as a note of interest on this class of emerging contaminants, and their relation to the reality of these compounds' bioactive properties should be regarded with caution.

Disclaimer:

The views expressed in this article are those of the authors and do not necessary represent the views of U.S. Environmental Protection Agency. The use of any named brands or products for this work does not constitute an official endorsement of those brands by the U.S. Environmental Protection Agency.

References

(1) Inoue K, Okada F, Ito R, Kato S, Sasaki S, Nakajima S, et al. Perfluorooctane sulfonate (PFOS) and related perfluorinated compounds in human maternal and cord blood samples: assessment of PFOS exposure in a susceptible population during pregnancy. Environmental health perspectives. 2004;112(11):1204-7: 0091-6765.

(2) Weiss JM, Andersson PL, Lamoree MH, Leonards PEG, van Leeuwen SPJ, Hamers T. Competitive binding of poly-and perfluorinated compounds to the thyroid hormone transport protein transthyretin. Toxicological sciences. 2009;109(2):206-16: 1096-6080.

(3) Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics. 2017;9(1):61.

(4) Ren X-M, Qin W-P, Cao L-Y, Zhang J, Yang Y, Wan B, et al. Binding interactions of perfluoroalkyl substances with thyroid hormone transport proteins and potential toxicological implications. Toxicology. 2016;366:32-42: 0300-483X.
(5) Yang X, Lyakurwa F, Xie H, Chen J, Li X, Qiao X, et al. Different binding mechanisms of neutral and anionic poly-/perfluorinated chemicals to human transthyretin revealed by In silico models. Chemosphere. 2017;182:574-83: 0045-6535.
(6) Zhang J, Kamstra JH, Ghorbanzadeh M, Weiss JM, Hamers T, Andersson PLJEs, et al. In silico approach to identify potential thyroid hormone disruptors among currently known dust contaminants and their metabolites. 2015;49(16): 10099-107