Journal of Hazardous Materials 492 (2025) 138069

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Molecular docking–QSAR–Kronecker-regularized least squares-based multiple machine learning for assessment and prediction of PFAS–protein binding interactions

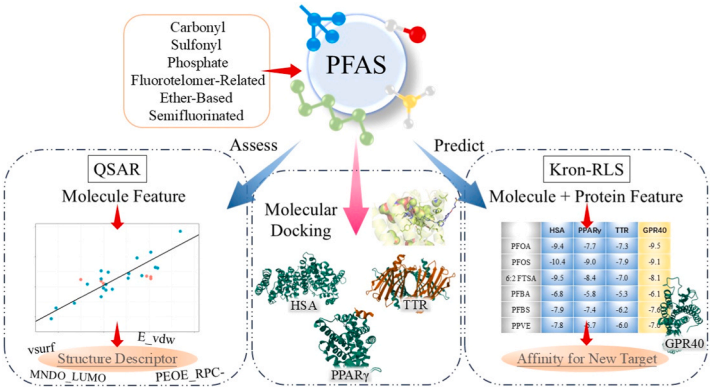
Lihui Zhao [a,b,1], Zixuan Zhang [a,1], Hailei Su [a], Wenjun Zhang [c], Jiaqi Sun [d], Yunxia Li [e], Miaomiao Teng [a,*]

[a] State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[b] College of geoexploration science and technology, Jilin University, Changchun 130026, China
[c] Key Laboratory of Integrated Regulation and Resources Development on Shallow Lakes of Ministry of Education, College of Environment, Hohai University, Nanjing 210098, China
[d] School of Energy and Environmental Engineering, University of Science and Technology Beijing, Beijing 100083, China
[e] College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

HIGHLIGHTS

- The binding pattern of 430 PFASs with HSA, PPARγ, TTR and GPR40 were calculated.
- Multiple machine learning models accelerated predictions of PFAS–protein interactions.
- Kron-RLS model is superior in predicting interactions with new targets.
- Study of PFAS–protein binding interactions helps to explain PFAS-mediated toxicity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
PFAS
Molecular docking
QSAR model
Kron-RLS model
Protein

ABSTRACT

Ubiquitous per- and poly-fluoroalkyl substances (PFAS) threaten human's health and attract worldwide attention. PFAS-mediated toxicity involves adverse effects of PFAS on proteins, and assessment of PFAS–protein binding interactions helps to explain PFAS' adverse effects on human health. In-silico modeling can generate information and decrease experimental costs. Accordingly, in this study, molecular docking was used to determine the binding affinities of 430 PFAS with human serum albumin (HSA), peroxisome proliferator-activated receptor gamma (PPARγ), and transthyretin (TTR). Specifically, analytic hierarchy process, fuzzy comprehensive evaluation, and quantitative structure–activity relationship model were used to assess and predict the

* Correspondence to: State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Anwai Dayang Fang 8#, Chaoyang District, Beijing 100012, China.
E-mail address: tengmiao0603@163.com (M. Teng).
[1] These authors contributed to the work equally and should be regarded as co-first authors.

https://doi.org/10.1016/j.jhazmat.2025.138069
Received 2 October 2024; Received in revised form 16 March 2025; Accepted 24 March 2025
Available online 29 March 2025

binding affinities between PFAS and HSA, PPARγ, and TTR. The binding patterns were determined by defining "PEOE_RPC-, E_vdw, MNDO_LUMO, and vsurf features" as key factors related to charge, energy and shape characteristic of PFAS. Finally, Kronecker-regularized least squares (Kron-RLS) model was applied to predict the binding affinities between PFAS– and G protein-coupled receptor 40 (GPR40), as a new target for prediction. Results showed that the Kron-RLS model exhibited good performance and generated precise predictions (R^2 = 0.94). In conclusion, this study demonstrated that computational simulations could be used to aid the scientific management of the growing number of PFAS, and could be broadened to include a wide range of environmental contaminations.

1. Introduction

Per- and poly-fluoroalkyl substances (PFAS) are a class of organic compounds that contain at least one fully fluorinated methyl or methylene carbon atom with no attached hydrogen, chlorine, bromine, or iodine atoms, and 5000 PFAS are known currently [1,2]. PFAS are used extensively but also pose environmental challenges. Perfluorooctanoic acid (PFOA), perfluorooctane sulfonate (PFOS), and their homologues are typical PFAS and are persistent and toxic hazards to the environment and humans [3,4]. However, there has been insufficient research on the potential adverse effects of emerging PFAS.

Advances in biological methods have allowed researchers to examine the molecular mechanisms of PFAS-mediated toxicity, and this has revealed that PFAS initiate biological events by binding to molecular targets [4]. PFAS can bind with proteins, such as transport proteins and nuclear proteins, in vivo [5–7]. For example, human serum albumin (HSA) carries PFAS to various tissues in circulating blood [8]; nuclear receptors such as the peroxisome proliferators activated receptor gamma (PPARγ) are activated by PFAS, leading to an increased risk of impaired bone development [9]; and PFAS compete with thyroid hormone (TH) for binding to transporter protein transthyretin (TTR), thereby influencing the TH homeostasis [10,11]. G protein-coupled receptor 40 (GPR40) is a long-chain fatty-acid receptor that is highly expressed in the pancreas [12]. Exposure to PFAS can have several adverse effects, e.g., it can stimulate insulin secretion, induce endoplasmic reticulum stress, and disrupt pancreatic organogenesis [13–15]. Thus, it is essential to study the interactions that triggers adverse outcomes.

Researchers have developed several strategies to characterize binding affinities and thus obtain protein–PFAS binding data. Degitz et al. [16] evaluated the bioactivity of 136 PFAS in the thyroid axis across nine in-vitro assays. They found that perfluoroheptanesulfonic acid, (perfluorobutyryl)-2-thenoylmethane, and perfluorohexanesulfonic acid showed the strongest inhibition to TTR (with EC_{50}s of 0.067, 0.183, and 0.184 μM, respectively), which were lower than that of thyroxine. Jia et al. [17] used microscale thermophoresis to examine the binding affinities between 11 PFAS and liver fatty acid-binding protein and has, respectively, and confirmed that the aforementioned PFAS and proteins exhibited ligand–protein interactions. Typical methods for exploring protein–ligand binding measure a range of dynamic equilibria and thermodynamic parameters and include equilibrium dialysis [18] and fluorescence competitive combination tests [19]. However, although these methods are intended to accomplish the same purpose, they use different characterization parameters, such as EC_{50}s, association constants, and dissociation constants. Thus, there is still a gap in the comparison among different studies, which needs more methods to unify valuable data.

Furthermore, although experimental data are valuable, experiments are expensive and time-consuming to perform and thus cannot meet practical needs. Consequently, researchers have used computational modeling methods, such as molecular docking, molecular dynamics simulations [20], and quantitative structure−activity relationship (QSAR) modeling [21], as more effective means of investigating binary complex interactions. Similarly, Li et al. [22] used a recombinant yeast-based assay and molecular docking to examine the interactions between 20 anthraquinones and estrogen receptor alpha, and revealed that there was a linear relationship between the observed and predicted results. Carlsson et al. [23] validated a sequence of docking, scoring, and molecular dynamics, and demonstrated that it enabled high throughput modeling. Docking can also be used to construct models to serve as representative data [24]. QSAR models are simple in-silico models that are used to establish a quantitative relationship between molecular structures and a specific endpoint. Hong et al. [25] used a QSAR model to predict the cytotoxicity of 29 nitrogenous disinfectant byproducts in human embryonic kidney cells. A well-trained QSAR model can also identify possible approaches for health risk assessment and greatly reduces the cost of experiments. Valid molecular descriptors ensure the feasibility of a QSAR model, as the modeling process consists of a series of steps, namely dataset preparation, molecular descriptor selection, regression model derivation, and model validation [26]. Selective two-dimensional (2D) descriptors, such as the highest or lowest occupied molecular orbital energies, dipole moments and octanol–water partition coefficients, are widely used as descriptors due to their comprehensibility and ease of calculation [25,27]. Recently, several studies screened the best descriptor combinations suitable for QSAR modeling from hundreds of global molecular descriptors containing various structural features of molecules [28–31]. The optimization of a QSAR model helps to reveal the essential structural features for building the model and to explain the key factors involved in a compound's induction of a given endpoints.

In many practical contexts, determining pollutant–protein interaction is not a simple binary classification problem (one that involves linking molecular structural features to biological endpoints), which is a core concept in most learning models. This means that a model can only precisely describe the interactions between a pollutant and a protein if it considers both species' structural features. Hence, we applied a Kronecker-regularized least squares (Kron-RLS) model for further machine learning. The Kron-RLS model was initially developed for predicting drug–target (protein) interactions [32,33]. It simultaneously generates a drug–drug similarity matrix and a protein–protein similarity matrix and uses computational shortcuts to rapidly train a model [34]. Yan et al. [35] used a Kron-RLS model to predict the associations between novel circular RNAs and it exhibited obtained good prediction performance. Van Laarhoven et al. [36] noted that a Kron-RLS model was sensitive to chemical similarity and exhibited reasonable precision if it was based on a large dataset. Their approach involved detecting new interactions from an established model. Given the above-mentioned advantages of a Kron-RLS model, we used such a model in a preliminary step to predict the binding affinities between PFAS and targeted proteins.

This study aims to evaluate the binding interactions between PFAS and several proteins and apply a new model for predicting binding affinities between PFAS and another protein. Considering the importance of physiological activity, we choose HSA, PPARγ, and TTR as studied proteins. We utilize a combination of molecular docking, machine learning methods including analytic hierarchy process (AHP), fuzzy comprehensive evaluation (FCE), and QSAR model to assess the comprehensive binding interactions. Subsequently, we introduce Kron-RLS model to predict the binding interactions between PFAS and GPR40, which is newly-applied in the field of environmental

contaminations evaluation. All software and algorithmic codes are accessible to ensure reproducibility of results. Overall, by examining how PFAS-mediated toxicity relies on interactions of PFAS with proteins, this study devised a method that could be applied to investigate other toxin–protein interactions.

2. Method

2.1. Molecules and proteins used in modeling

A dataset containing 430 PFAS was used in this study as listed in Table S1, available for download at (https://comptox.epa.gov/dashboard/chemical_lists/EPAPFASINV). The dataset was adequately represented the diversity of the commercial PFAS [37]. HSA (Protein Data Bank (PDB) id: 1E7G), PPARγ (PDB id: 3U9Q), TTR (PDB id: 2F7I), and GPR40 (PDB id: 4PHU) were the studied proteins. Two-dimensional structures of PFAS were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/), and three-dimensional structures and sequences of proteins were downloaded from the PDB (https://www.rcsb.org/).

2.2. Molecular docking and verification

Molecular docking, scoring and ranking process were accomplished using AutoDock vina software. AutoDock vina is an efficient and accurate tool for mass molecular docking work [38]. PFAS were set as ligands, and HSA, PPARγ, TTR, and GPR40 were set as proteins (i.e., receptors). Flexible docking strategy, which allows diversified conformations of both ligand and receptor in a certain space, was applied to predict the binding affinity between PFAS and receptors [39]. The docking grid was located at the center of each active site and the box size was 26 Å that covered the whole active site. Specifically, the grid coordinate of PPARγ was x = 1.465, y = 1.199 and z = 21.819; the grid coordinate of TTR was x = 22.494, y = 82.045 and z = 8.291; and the grid coordinate of GPR40 was x = -41.704, y = -2.588 and z = 60.604. HSA is a protein with couple of activate sites, and eight grid boxes were taken into account in this study. The grid coordinates of were x = -1.214, y = 4.148 and z = 39.286; x = 11.889, y = 10.431 and z = 14.193; x = 10.076, y = 4.110 and z = 19.817; x = 23.744, y = 5.663 and z = -1.731; x = 35.563, y = 15.558 and z = 34.937; x = 33.231, y = 14.498 and z = 7.791; x = 47.731, y = 14.448 and z = 18.111; and x = 54.050, y = 8.238 and z = 21.460. On the basis of the simulation results, twenty ligand–receptor complexes were generated. The cluster analysis module of AutoDock vina was used to assign points for molecular docking energy, and the complex with lowest-energy was chosen for further study.

To further explore the difference binding affinities among PFAS categories, PFAS were divided into eight groups as referred to Buck, Korzeniowski, Laganis and Adamsky [40].

2.3. Comprehensive assessment of interactions between PFAS and proteins

A comprehensive assessment of interactions between compounds and proteins aims to holistically appraise a compound–protein system according to a multi-attribute architecture [41]. In the current study, the comprehensive binding affinity between PFAS and proteins was determined by the integration of an AHP and an FCE. AHP is a popular technique used to support decision making, and uses pairwise comparisons for systematizing and structuring decision-making [42,43]. First, a pairwise comparison matrix was built by estimating binding affinity on a scale from 1 to 5, corresponding to "equal binding affinity" and "much stronger binding affinity", respectively. Next, an arithmetic method was used to calculate the weight vector for each protein. Finally, a consistency test was conducted and the random consistency index was required to be less than 0.1.

FCE applies fuzzy mathematics to describe objects restricted by various factors, with which qualitative results can be transformed into quantitative results [44]. In the current study, each docking score was fuzzily determined on a scale from 1 to 10, indicating low binding affinity to strong binding affinity, respectively, and a single-factor evaluation set r_i was built by matching docking scores (Eq. (1)). Subsequently, a comprehensive evaluation matrix R was generated based entirely on fuzzy estimation (Eq. (2)). Finally, a comprehensive index (CI) was calculated based on a weight vectors matrix W via an AHP and used to represent the comprehensive interaction (Eq. (3)).

$$r_i = \{r_{i1}, r_{i2}, \cdots, r_{im}\} \quad (1)$$

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1j} \\ \vdots & \ddots & \vdots \\ r_{i1} & \cdots & r_{ij} \end{bmatrix} \quad (2)$$

$$CI = R * W \quad (3)$$

2.4. Descriptor screening, QSAR modeling, and validation

First, the robustness and predictive ability of the QSAR model were demonstrated. The PFAS, excluded twelve PFAS that failed to bind and seven PFAS that hardly to bind according to molecular docking, were divided into an internal training set and an external validation set in an approximate 7:3 ratio. Specifically, 287 PFAS comprised the validation set, and the remaining 124 PFAS comprised the training set.

Next, after minimizing the structures of each PFAS in the training set (under MMFF94 force field), 206 2D descriptors and 148 3D descriptors were calculated for each PFAS in this set using Molecular Operation Environment software. The zero value and constant descriptors were excluded to prevent the generation of meaningless data, and a correlation analysis was conducted to remove the redundant descriptors (those with an $R > 0.95$) [45]. Consequently, 72 descriptors were retained for QSAR modeling.

The least absolute shrinkage and selection operator (LASSO) combined with 10-fold-cross-validation was used for variable selection to prevent multicollinearity and filter out valuable information. In the last step, multiple linear regression (MLR) was used for modeling, as this method effectively establishes relationships between multiple molecular descriptors and biological activities. As it is important to determine an adjusted coefficient of determination (R^2) to calibrate a QSAR model, a residue plot was constructed for internal verification. In addition, a set of test data was used to examine the generalizability of the model.

To analysis the role of molecular descriptors in protein–PFAS interaction, Shapley Additive explanation (SHAP) was introduced to interpret models. The SHAP approach is a useful tool for prioritization of features that determine the activity prediction [46].

2.5. Kron-RLS modeling and prediction

Three different kernels were demanded to build protein–PFAS Kron-RLS model. To improve the generalization capacity of Kron-RLS model, 418 PFAS were included in the process of modeling. PFAS structure similarity matrix (K_m) served as the chemical kernel and was determined using the PubChem Score Matrix Service (https://pubchem.ncbi.nlm.nih.gov/docs/score-matrix-service) based on Tanimoto similarity. A protein structure similarity matrix (K_p) served as the protein kernel and was determined from the Smith–Waterman score based on the blocks substitution 62 matrix. The sequences of binding pockets were used in calculations to obtain effective descriptions of binding interactions. Specifically, the sequence of a binding pocket was defined as a continuous segment on a protein containing all of the pocket residues, which were defined as the residues encircling the ligand within a distance of 5 Å [47]. A protein–PFAS docking score matrix was used as a chemical–protein pair kernel (K_{m-p}). The data of PFAS bound to HSA, PPARγ, and TTR, respectively, were used to train the Kron-RLS model. $K_{m\ (418,418)}$, $K_{p\ (3,3)}$ (HSA, PPARγ, and TTR) and their related matrix $K_{m-p\ (418,3)}$

were used as training set optimization functions to build the Kron-RLS model. During modeling, the data of the training set were disordered, and the best regularization parameter λ for the Kron-RLS model was obtained. The concordance index (cindex) was calculated by a leave-one-out process for screening λ and evaluating the training model. As a model for prediction, the cindex greater than 0.5 was eligible for predictive power. After adjustment, the external data of GPR40 bound to each PFAS was input for prediction of GPR40–PFAS binding affinities. $K_{m\ (418,418)}$, $K_{p\ (4,4)}$ (for HSA, PPARγ, TTR and GPR40) and their related matrix $K_{m-p\ (318,4)}$ (including unknown data) were used for prediction in the Kron-RLS model. The binding affinities of PFAS bound to GPR40 that obtained by molecular docking, as the validation set, were utilized to evaluate the prediction results. The devised algorithms were generated using RLScore software (https://github.com/aatapa/RLScore).

3. Results and discussion

3.1. Assessment of affinities between PFAS and proteins

Molecular docking was performed to determine the interactions and binding affinities between PFAS and HSA, PPARγ, and TTR, respectively. HSA, PPARγ, and TTR were used because of their broad range of physiological activities. HSA is a crucial protein in serum and binds PFAS with high affinity [18,48]. Epidemiologic studies have revealed that PFAS are associated with increases in body mass index, leptin concentrations, and incident diabetes [49–51]. Moreover, perturbations of lipid and glucose metabolism may be mediated by the effects of PFAS on PPARγ [52]. In addition, PFAS exposure is associated with thyroid system dysfunction, which adversely affects human health and development [53,54].

During molecular docking, the final binding phase of each molecule was automatically selected according to the maximum clustering phase and optimal energy score. There are approximately 10 binding sites on HSA that are capable of binding PFAS [55], whereas there is only one binding site each on PPARγ, TTR and GPR40. Thus, the above-mentioned automatic selection enhances model efficiency and obtains reliable results for given objective [7,56]. The molecular docking results are shown in Fig. 1 and Table S2, with low values indicating strong binding.

A hierarchical model was constructed to comprehensively determine the binding affinities between the PFAS and HSA, PPARγ, and TTR, respectively. With reference to AHP [44], the order of binding affinities between the PFAS and the proteins was HSA > PPARγ > TTR, and the weight vector was judged via the pairwise comparison matrix (Table S3). Therefore, the weight vector was (0.581, 0.309, 0.110) (Eq. (4)). A consistency test revealed that the random consistency index was adequate, as it equaled 0.0036 < 0.1 (Eqs. (5–8)).

$$A=\begin{Bmatrix}1 & 2 & 5\\ \frac{1}{2} & 1 & 3\\ \frac{1}{5} & \frac{1}{3} & 1\end{Bmatrix}\rightarrow\begin{Bmatrix}\frac{10}{17} & \frac{3}{5} & \frac{5}{9}\\ \frac{5}{17} & \frac{3}{10} & \frac{1}{3}\\ \frac{2}{17} & \frac{1}{10} & \frac{1}{9}\end{Bmatrix}\rightarrow\begin{Bmatrix}0.581\\ 0.309\\ 0.110\end{Bmatrix}=W \tag{4}$$

$$AW=\begin{Bmatrix}1 & 2 & 5\\ \frac{1}{2} & 1 & 3\\ \frac{1}{5} & \frac{1}{3} & 1\end{Bmatrix}*\begin{Bmatrix}0.581\\ 0.309\\ 0.110\end{Bmatrix}=\begin{bmatrix}1.747\\ 0.929\\ 0.329\end{bmatrix} \tag{5}$$

$$\lambda=\frac{1}{n}\sum_{i=1}^{n}\frac{[AW]_i}{W_i}=3.004 \tag{6}$$

$$CI=\frac{\lambda-n}{n-1}=0.0018 \tag{7}$$

$$CR=\frac{CI}{RI}=0.0036<0.1 \tag{8}$$

Moreover, the intensities of binding affinities were determined by FCE. Table S4 show the scoring system, and the hierarchical division was

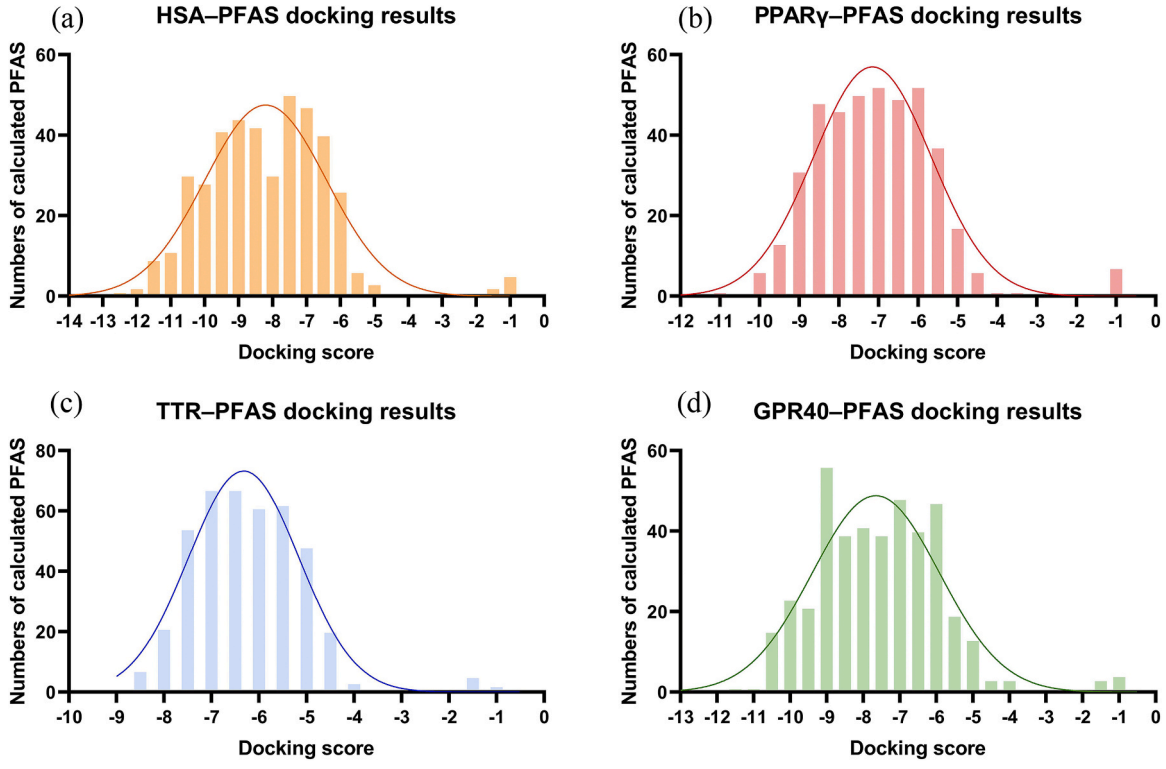


Fig. 1. Statistical analysis of (a) HSA, (b) PPARγ, (c) TTR and (d) GPR40 –PFAS molecular docking data.

equidistant such that close docking scores were classified into the same level. Based on the weight vector, the CI of 411 PFAS were calculated and are shown in Fig. 2 and Table S5, where a high value indicates a strong binding affinity. The index describes the binding affinities between the PFAS and HSA, PPARγ, and TTR.

3.2. QSAR modeling of key interactions between PFAS and proteins

The whole descriptor screening method was used to build a QSAR model to explore the key structure descriptors of protein–PFAS binding interactions. Compared with the typical method used to select structure descriptors, this method provides more multidimensional insights based on the extensive analysis of the physical and chemical properties of molecules.

During the LASSO regression, the descriptors subset was obtained when lambda.min = 0.009576 (Figure S1). Next, the descriptors subset was also validated by a variance inflation factor (VIF) analysis. VIFs are often used to represent the extent of collinearity in a multiple linear regression model, and in a QSAR model, VIFs of less than 10 indicate that the descriptors are well correlated [57]. The selected structure descriptors for QSAR modeling were listed in Table S6, and correlation test in Figure S2 provided visualization.

The MLR method is suitable for establishing a QSAR model based on data in which the independent variables are less than the response variables. Fig. 3a shows that the CIs and predicted results fitted well into the QSAR model, as indicated by an adjusted $R^2 = 0.89$ and p-value < 0.001. The QSAR model of the key structure descriptors describing the binding of PFAS to proteins was obtained as follows:

According to the Tukey–Anscombe plot (residual vs. fitted), there was an unordered pattern satisfying the homogeneity hypothesis of variance. The quantile–quantile residues plot demonstrated that the points were reasonably well distributed (Figure S3). Thus, the model well explained the binding affinities between the PFAS and the proteins. Recent studies have emphasized the necessity of externally validating data, aside from using a training set [58–60]. According to Golbraikh and Tropsha [61] criteria, a well-performing model should have an R almost equal to 1 and an R^2 equal to either R_0^2 or $R_0'^2$, and the slope of the corresponding regression lines (k or k′) should equal 1. In this study, an external test set (Section 2.4) was used to test the robustness and predictive ability of the QSAR model. The calculated $R = 0.903$, $(R_0'^2 - R_0^2)/R^2 = 0.0057$ (i.e., < 0.100), $k = 1.027$, and $k' = 0.954$. The results of QSAR model revealed that the PFAS with phosphate acid group showed precisely prediction among the PFAS categories while the PFAS with sulfonic acid group were a bit bias. Overall, the internal and external validations confirm that the QSAR model well described the respective binding affinities between the PFAS and HSA, PPARγ, and TTR, and the reliability of molecular docking was also verified.

The selected in-silico molecular descriptors contained crucial information on the structure feature of PFAS. These descriptors were strongly correlated with the structural information of PFAS, and they did not exhibit multicollinearity in models (Figure S2, Table S5). To further determine the role of each molecular descriptor in the binding patterns of the PFAS with proteins, a SHAP analysis was built to interpret the QSAR model. According to Fig. 3b, PEOE_RPC-, E_vdw and MNDO_-LUMO were identified as the most important features, followed by BCUT_PEOE_0, vsurf_CW3, and etc. PEOE_RPC- represent the molecules' relative negative charge, thereby suggesting whether electrostatic interactions are involved in the molecules' binding with targets [62]. E_vdw reflects the intrinsic vdW properties of calculated PFAS [63]. MNDO_LUMO indicates the molecule's abilitiy to accept electrons. Vsurf are a series of descriptors describe the interaction of molecules with hydrophobic and hydrophilic part of the protein, which was previously validated in a pharmacokinetics context [64,65]. The amphiphilic moment (vsurf_A) and molecular shape complexity (vsurf_ID2) implied direct implications for the steric influences. According to these analyses, the surface charge distribution and hydrophilicity/hydrophobicity of PFAS were the primary contributors to their binding interactions with the proteins. The results indicate that PFAS with sulfonic acid groups exhibited stronger protein binding abilities than those with carboxylic acid groups, which is consistent with the findings of previous studies [6, 66]. Moreover, the length of the fluorocarbon chains affected hydrophobicity, such that the longer the carbon chain of a PFAS, the stronger was its protein-binding ability.

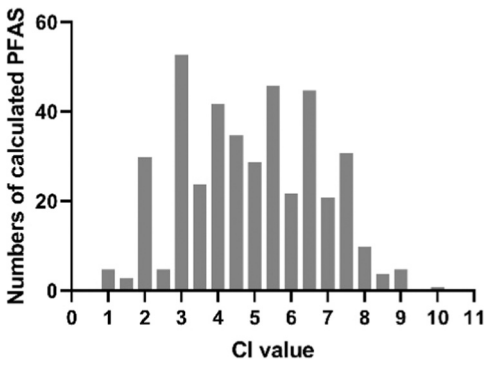


Fig. 2. Distribution of CI values to be representative of binding affinities between the PFAS and the proteins.

LASSO regression screening reveals the most effective subset of descriptors that affect the predicted endpoint [67] and was used to comprehensively characterize the binding affinities between the PFAS and the proteins in the current study. Commonly used or key descriptors may be neglected or not exist in a final subset obtained via in-silico screening of hundreds of descriptors [68,69]. Therefore, an analysis of the key factors that affect binding affinities should both consider screened descriptors and their implications.

3.3. Kron-RLS model-based prediction of interactions between the PFAS and GPR40

The QSAR model only focused on the relationship between the structural characteristics of PFAS and their biological activities, i.e., the binding affinities between PFAS and proteins, given the differences between proteins. Therefore, we also used a Kron-RLS model to obtain a broader understanding of the above-mentioned relationship. The structural features of PFAS, sequences of three well-studied proteins (HSA, PPARγ, and TTR) and paired binding affinities were used to train a Kron-RLS model. A previous study indicated that small molecules were conservative ligands for proteins with sequence identities of over 80 % [70]. Moreover, proteins with similar sequences usually perform analogous functions, and thus sequence-based strategies are often used for target identification [71]. Chen et al. [47] evaluated the prediction ability of a model using three kinds of protein sources (e.g., entire protein sequences, pocket sequences, and pocket residues). They found that input of pocket residues and pocket sequences, respectively, gave approximately similar and reasonable results (identity area under the curve = 90.0 % and 86.5 %, respectively), whereas input of an entire protein sequence generated much worse results (identity area under the curve = 56.9 %). Thus, in the current study, the information of binding pockets was applied in Kron-RLS modeling to enhance its performance. The binding affinities between the PFAS and GPR40 were predicted to test the scalability of the Kron-RLS model.

Kernels K_m, K_p, and K_{m-p} were calculated using the methods described in Section 2.5. First, the optimal regularization coefficient λ of the two kernels (K_m and K_p) were screened, as λ is a discretionary parameter for balancing prediction error with model complexity [32]. Accordingly, for every λ value there is a cindex, which represents the best description of the features in a training set $K_{m-p\ (27,3)}$. As shown in Fig. 4, when λ was equal to 2^3, the cindex was optimal at a value of 0.9130, indicating the best fitting parameter. Thus, this λ was taken as

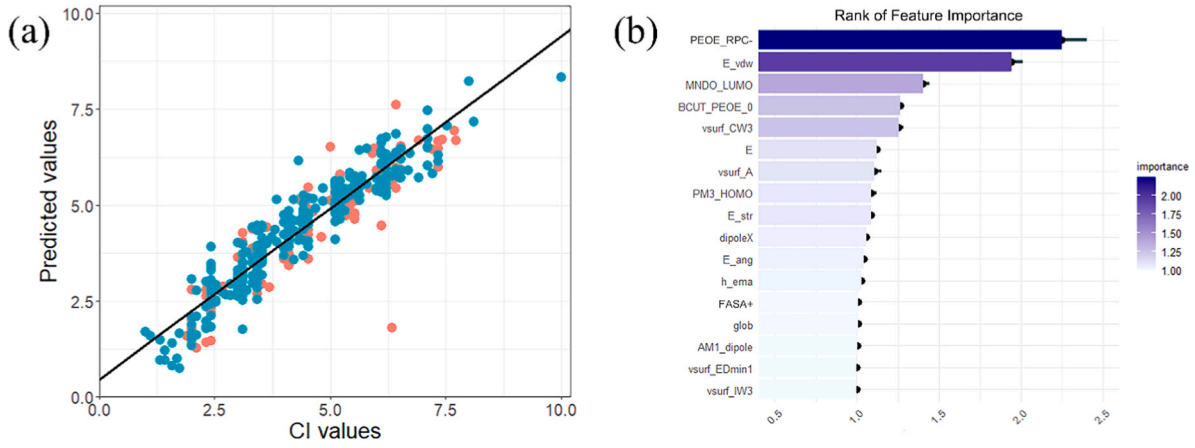


Fig. 3. QSAR model of comprehensive binding affinities between PFAS and HSA, PPARγ, and TTR. (a) Scatter plot of QSAR model. The green points correspond to the training set and the red points correspond to the validation set; (b) SHAP analysis of QSAR model.

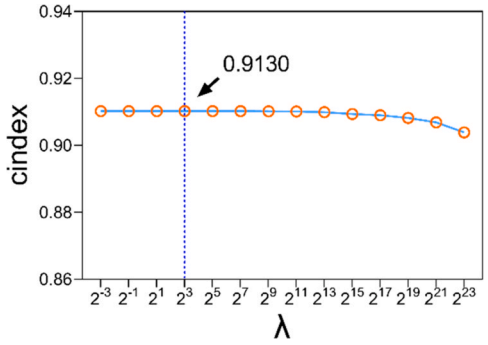


Fig. 4. Regularization coefficient optimization curve of the Kron-RLS model.

the regularization coefficient of the final model. A linear kernel for both input domains was used in the regression approach, as it operated well on small matrices. After modeling, the binding affinities between the PFAS and GPR40 were predicted. The R^2 of Kron-RLS model reached 0.94.

The Kron-RLS model was trained on molecular docking results and thus its predictions exhibited a similar trend to these results, i.e., the lower the value of a prediction, the higher the binding affinity. Generally, the predictions demonstrated that there were strong binding affinities between the PFAS and GPR40. Moreover, the experimental data of GPR40–PFAS and TTR–PFAS complexes were of the same order of magnitude, which confirmed the validity of the predictions [72]. According to Fig. 5a, most PFAS in eight categories performed binding capacity to GPR40. For a given fluorinated-carbon chain length, the binding affinities between the PFAS with a sulfonic acid group and proteins were stronger than those between the PFAS with a carboxylic acid group and proteins. Interestingly, PFAS with phosphate acid groups seemed to be stronger binding affinities (which was underestimate previously), however, insufficient data limited the observation. These results are consistent with the findings of Qin et al. [13], suggesting that positive predictions are generated for active ligands.

Zhang et al. [73] reported that PFOS stimulates insulin secretion via the GPR40 pathway. Hoyeck et al. [74] also suggested that long-chain PFAS may activate GPR40 to cause β-cell toxicity. In the current study, as expected, PFAS containing 8–12 carbon atoms (e.g., PFOA, PFOS, PFNS, PFDoS, and PFUnA) bound GRP40 strongly (Table S2). However, the Kron-RLS model was not effective in predicting inactive ligands (e.g., PFAS with short carbon chains and complex structures). To verify this result, a detailed molecular docking analysis was performed to assess the binding patterns of the PFAS with GPR40. As shown in Figs. 5b and 5c, this analysis revealed that PFAS were bound to the GPR40 binding pocket. For example, the binding pocket on GPR40 was occupied fully by PFOS. PFOS formed hydrogen bonds with R258, R183 and Y240 of GPR40 through sulfonic acid group, as well as some hydrophobic and polar interactions around the carbon chain. Given the physiological function of GPR40, the results indicate that the adverse effects of PFAS on GPR40 and insulin secretion have previously been substantially underestimated. More research is needed to explore the adverse effects of PFAS on these targets.

In comparison to QSAR model, the Kron-RLS model was superior in predicting untested interactions between molecules and new proteins (with a reported binding pocket). More complicated models, such as graph convolutional networks and graph attention networks, can also be used for this purpose [75,76], but compared with the Kron-RLS model, they are more difficult and costly to implement and require many more data. Therefore, the Kron-RLS model could be used for the rapid and accurate generation of predictions of the adverse effects of pollutants on other protein targets. It is worth noting that Kron-RLS model based on flexible docking data is much better than semi-flexible docking data (data not shown), and it might be explained that the molecular docking approach containing complex conformations of ligands and receptors can accurately match the underlying information of binding pocket residues as input of the Kron-RLS model. It should be highlighted that the better application prospect is to build Kron-RLS model via reliable experimental data rather than stimulated data. In this regard, a first step is to obtain sufficient data to generate a precise model. In addition, a model can be improved by increasing the accuracy of the protein structure similarity algorithm that is used. In future, the modeling process of the Kron-RLS model can be further optimized by using common feature descriptors instead of simple molecular fingerprints to establish a structural similarity matrix.

4. Conclusion

Overall, the discovery and commercialization of PFAS have increased convenience in our lives, but PFAS nevertheless have adverse effects on humans and the environment. Therefore, PFAS-mediated toxicities, such as PFAS' interactions with proteins, should be examined. In this study, we obtained molecular docking predictions, realizing an abundant dataset for research. Our QSAR model confirmed that molecular descriptors of PFAS (PEOE_RPC-, E_vdw, MNDO_LUMO, and vsurf features) are key factors that describe the PFAS' pattern of binding to HSA, PPARγ, and TTR. In addition, the molecular descriptor screening method helped to reveal crucial information. Finally, a Kron-RLS model was established to predict the binding affinities between PFAS and GPR40. The validated Kron-RLS model showed the expected prediction

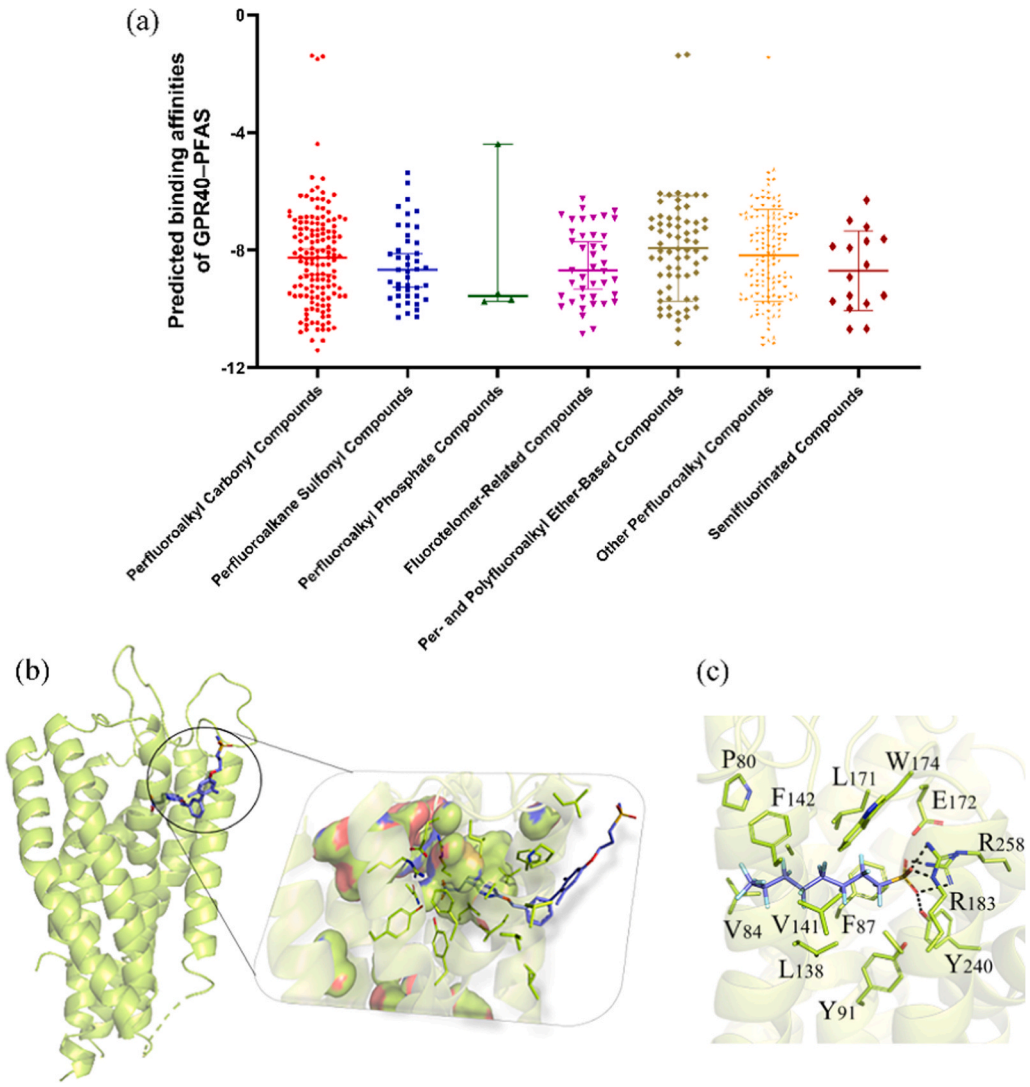


Fig. 5. (a) Comparison of predicted binding affinities from different PFAS categories calculated by Kron-RLS model. Scatter plot reported mean values with 95 % confidence intervals; (b) Binding pocket (active site) on GPR40; (c) Binding pattern analysis of GPR40–PFOS.

ability on another target during model training. The results suggest that PFAS have adverse effects on human health and that the risks of the growing numbers of PFAS could be managed by use of the Kron-RLS model. Furthermore, the model could be broadened to allow its application in predicting other chemicals' abilities to contaminate the environment.

Environmental Implication

PFAS are a class of environmental contaminants that arouse great concern because of environment risks. The study of protein–PFAS binding interactions can explain the biological mechanisms of toxicology effect. In this study, multiple machine learning models were successfully applied to predict the binding affinities of PFASs and important proteins including HSA, PPARγ, TTR and GPR40. It is conceivable that in-silico modeling could save research costs and accelerate scientific assessments on PFAS, which powerfully contributes to the theme of environment and human health.

CRediT authorship contribution statement

Li Yunxia: Investigation. **Teng Miaomiao:** Writing – review & editing, Supervision. **Zhang Zixuan:** Investigation. **Su Hailei:** Writing – review & editing. **Zhao Lihui:** Writing – original draft, Methodology, Investigation. **Zhang Wenjun:** Writing – review & editing. **Sun Jiaqi:** Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42207335), and Young Elite Scientists Sponsorship Program by CAST (2022QNRC001).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2025.138069.

Data Availability

Data will be made available on request.

References

[1] OECD, 2021. Reconciling terminology of the universe of per- and polyfluoroalkyl substances: recommendations and practical guidance. Ser Risk Manag No 61.
[2] OECD, 2018. Toward a new comprehensive global database of per- and polyfluoroalkey substances (PFASs): summary report on updating the OECD 2007 list of per- and polyfluoroalkey substances (PFASs). Ser Risk Manag No 39.
[3] Podder, A., Sadmani, A.H.M.A., Reinhart, D., Chang, N.-B., Goel, R., 2021. Per and poly-fluoroalkyl substances (PFAS) as a contaminant of emerging concern in surface water: a transboundary review of their occurrences and toxicity effects. J Hazard Mater 419, 126361.
[4] Fenton, S.E., Ducatman, A., Boobis, A., DeWitt, J.C., Lau, C., Ng, C., Smith, J.S., Roberts, S.M., 2021. Per- and Polyfluoroalkyl substance toxicity and human health review: current state of knowledge and strategies for informing future research. Environ Toxicol Chem 40, 606–630.
[5] Forsthuber, M., Kaiser, A.M., Granitzer, S., Hassl, I., Hengstschlager, M., Stangl, H., Gundacker, C., 2020. Albumin is the major carrier protein for PFOS, PFOA, PFHxS, PFNA and PFDA in human plasma. Environ Int 137, 105324.
[6] Zhao, L., Teng, M., Zhao, X., Li, Y., Sun, J., Zhao, W., Ruan, Y., Leung, K.M.Y., Wu, F., 2023. Insight into the binding model of per- and polyfluoroalkyl substances to proteins and membranes. Environ Int 175, 107951.
[7] Søderstrøm, S., Lille-Langøy, R., Yadetie, F., Rauch, M., Milinski, A., Dejaegere, A., Stote, R.H., Goksøyr, A., Karlsen, O.A., 2022. Agonistic and potentiating effects of perfluoroalkyl substances (PFAS) on the Atlantic cod (Gadus morhua) peroxisome proliferator-activated receptors (Ppars). Environ Int 163, 107203.
[8] Gao, Y., Fu, J., Cao, H., Wang, Y., Zhang, A., Liang, Y., Wang, T., Zhao, C., Jiang, G., 2015. Differential accumulation and elimination behavior of perfluoroalkyl Acid isomers in occupational workers in a manufactory in China. Environ Sci Technol 49, 6953–6962.
[9] Buckley Jessie, P., Kuiper Jordan, R., Lanphear Bruce, P., Calafat Antonia, M., Cecil Kim, M., Chen, A., Xu, Y., Yolton, K., Kalkwarf Heidi, J., Braun Joseph, M., 2021. Associations of maternal serum perfluoroalkyl substances concentrations with early adolescent bone mineral content and density: the health outcomes and measures of the environment (HOME) study. Environ Health Perspect 129, 097011.
[10] Berg, V., Nøst, T.H., Hansen, S., Elverland, A., Veyhe, A.-S., Jorde, R., Odland, J.Ø., Sandanger, T.M., 2015. Assessing the relationship between perfluoroalkyl substances, thyroid hormones and binding proteins in pregnant women; a longitudinal mixed effects approach. Environ Int 77, 63–69.
[11] Ren, X.-M., Qin, W.-P., Cao, L.-Y., Zhang, J., Yang, Y., Wan, B., Guo, L.-H., 2016. Binding interactions of perfluoroalkyl substances with thyroid hormone transport proteins and potential toxicological implications. Toxicology 366, 32–42.
[12] Itoh, Y., Kawamata, Y., Harada, M., Kobayashi, M., Fujii, R., Fukusumi, S., Ogi, K., Hosoya, M., Tanaka, Y., Uejima, H., Tanaka, H., Maruyama, M., Satoh, R., Okubo, S., Kizawa, H., Komatsu, H., Matsumura, F., Noguchi, Y., Shinohara, T., Hinuma, S., Fujisawa, Y., Fujino, M., 2003. Free fatty acids regulate insulin secretion from pancreatic β cells through GPR40. Nature 422, 173–176.
[13] Qin, W.-P., Cao, L.-Y., Li, C.-H., Guo, L.-H., Colbourne, J., Ren, X.-M., 2020. Perfluoroalkyl substances stimulate insulin secretion by islet β cells via G protein-coupled receptor 40. Environ Sci Technol 54, 3428–3436.
[14] Sant, K.E., Venezia, O.L., Sinno, P.P., Timme-Laragy, A.R., 2019. Perfluorobutanesulfonic acid disrupts pancreatic organogenesis and regulation of lipid metabolism in the Zebrafish, Danio rerio. Toxicol Sci 167, 258–268.
[15] Zhao, L., Teng, M., Shi, D., Sun, J., Li, Y., Zhang, Z., Zhu, W., Wu, F., 2024. Adverse impacts of environmentally relevant PFOS alternatives on mice pancreatic tissues. Sci Total Environ 909, 168649.
[16] Degitz, S.J., Olker, J.H., Jeffrey, Denny, S., Degoey, P.P., Hartig, P.C., Cardon, M. C., Eytcheson, S.A., Haselman, J.T., Mayasich, S.A., Hornung, M.W., 2023. In vitro screening of per- and polyfluorinated substances (PFAS) for interference with seven thyroid hormone system targets across nine assays. Toxicol Vitr, 105762.
[17] Jia, Y., Zhu, Y., Xu, D., Feng, X., Yu, X., Shan, G., Zhu, L., 2022. Insights into the competitive mechanisms of per- and polyfluoroalkyl substances partition in liver and blood. Environ Sci Technol 56, 6192–6200.
[18] Gao, K., Zhuang, T., Liu, X., Fu, J., Zhang, J., Fu, J., Wang, L., Zhang, A., Liang, Y., Song, M., Jiang, G., 2019. Prenatal exposure to per- and polyfluoroalkyl substances (PFASs) and association between the placental transfer efficiencies and dissociation constant of serum proteins–PFAS complexes. Environ Sci Technol 53, 6529–6538.
[19] Jackson, T.W., Scheibly, C.M., Polera, M.E., Belcher, S.M., 2021. Rapid characterization of human serum albumin binding for per- and polyfluoroalkyl substances using differential scanning fluorimetry. Environ Sci Technol 55, 12291–12301.
[20] Delva-Wiley, J., Jahan, I., Newman, R.H., Zhang, L., Dong, M., 2021. Computational analysis of the binding mechanism of GenX and HSA. ACS Omega 6, 29166–29170.
[21] Cao, H., Zhou, Z., Wang, L., Liu, G., Sun, Y., Wang, Y., Wang, T., Liang, Y., 2019. Screening of potential PFOS alternatives to decrease liver bioaccumulation: experimental and computational approaches. Environ Sci Technol 53, 2811–2819.
[22] Li, F., Li, X., Shao, J., Chi, P., Chen, J., Wang, Z., 2010. Estrogenic activity of anthraquinone derivatives: *in vitro* and in silico studies. Chem Res Toxicol 23, 1349–1355.
[23] Carlsson, J., Boukharta, L., Åqvist, J., 2008. Combining docking, molecular dynamics and the linear interaction energy method to predictbinding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J Med Chem 51, 2648–2656.
[24] Li, X., Gu, W., Zhang, B., Xin, X., Kang, Q., Yang, M., Chen, B., Li, Y., 2022. Insights into toxicity of polychlorinated naphthalenes to multiple human endocrine receptors: mechanism and health risk analysis. Environ Int 165, 107291.
[25] Hong, H., Lu, Y., Zhu, X., Wu, Q., Jin, L., Jin, Z., Wei, X., Ma, G., Yu, H., 2023. Cytotoxicity of nitrogenous disinfection byproducts: A combined experimental and computational study. Sci Total Environ 856, 159273.
[26] Myint, K.Z., Xie, X.-Q., 2010. Recent Advances in Fragment-Based QSAR and Multi-Dimensional QSAR Methods. Int J Mol Sci 11, 3846–3866.
[27] Xiao, R., Ye, T., Wei, Z., Luo, S., Yang, Z., Spinney, R., 2015. Quantitative Structure–Activity Relationship (QSAR) for the Oxidation of Trace Organic Contaminants by Sulfate Radical. Environ Sci Technol 49, 13394–13402.
[28] Hoover, G., Kar, S., Guffey, S., Leszczynski, J., Sepúlveda, M.S., 2019. In vitro and in silico modeling of perfluoroalkyl substances mixture toxicity in an amphibian fibroblast cell line. Chemosphere 233, 25–33.
[29] Jean, J., Kar, S., Leszczynski, J., 2018. QSAR modeling of adipose/blood partition coefficients of Alcohols, PCBs, PBDEs, PCDDs and PAHs: A data gap filling approach. Environ Int 121, 1193–1203.
[30] García-González, L.A., Marrero-Ponce, Y., Brizuela, C.A., García-Jacas, C.R., 2023. Overproduce and select, or determine optimal molecular descriptor subset via configuration space optimization? Application to the prediction of ecotoxicological endpoints. Mol Inform 42, 2200227.
[31] Motamedi, F., Pérez-Sánchez, H., Mehridehnavi, A., Fassihi, A., Ghasemi, F., 2022. Accelerating Big Data Analysis through LASSO-Random Forest Algorithm in QSAR Studies. Bioinformatics 38, 469–475.
[32] Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., Aittokallio, T., 2014. Toward more realistic drug–target interaction predictions. Brief Bioinforma 16, 325–337.
[33] Ben-Hur, A., Noble, W.S., 2005. Kernel methods for predicting protein–protein interactions. Bioinformatics 21, i38–i46.
[34] Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J., Zhang, Y., 2015. Drug–target interaction prediction: databases, web servers and computational models. Brief Bioinforma 17, 696–712.
[35] Yan, C., Wang, J., Wu, F.-X., 2018. DWNN-RLS: regularized least squares method for predicting circRNA-disease associations. BMC Bioinforma 19, 520.
[36] Van Laarhoven, T., Nabuurs, S.B., Marchiori, E., 2011. Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics 27, 3036–3043.
[37] Patlewicz, G., Richard Ann, M., Williams Antony, J., Grulke Christopher, M., Sams, R., Lambert, J., Noyes Pamela, D., DeVito Michael, J., Hines Ronald, N., Strynar, M., Guiseppi-Elie, A., Thomas Russell, S., 2019. A Chemical Category-Based Prioritization Approach for Selecting 75 Per- and Polyfluoroalkyl Substances (PFAS) for Tiered Toxicity and Toxicokinetic Testing. Environ Health Perspect 127, 014501.
[38] 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, 455–461.
[39] Feng, C., Lin, Y., Le, S., Ji, J., Chen, Y., Wang, G., Xiao, P., Zhao, Y., Lu, D., 2024. Suspect, Nontarget Screening, and Toxicity Prediction of Per- and Polyfluoroalkyl Substances in the Landfill Leachate. Environ Sci Technol 58, 4737–4750.
[40] Buck, R.C., Korzeniowski, S.H., Laganis, E., Adamsky, F., 2021. Identification and classification of commercially relevant per- and poly-fluoroalkyl substances (PFAS). Integr Environ Assess Manag 17, 1045–1055.
[41] Ren, Z., Wang, Y., Xu, H., Li, Y., Han, S., 2019. Fuzzy Comprehensive Evaluation Assistant 3D-QSAR of Environmentally Friendly FQs to Reduce ADRs. Int J Environ Res Public Health 16, 3161.
[42] Ho, W., Ma, X., 2018. The state-of-the-art integrations and applications of the analytic hierarchy process. Eur J Oper Res 267, 399–414.
[43] Zhang, C., Chen, M., 2018. Prioritising alternatives for sustainable end-of-life vehicle disassembly in China using AHP methodology. Technol Anal Strateg Manag 30, 556–568.
[44] Zhang, P., Feng, G., 2018. Application of fuzzy comprehensive evaluation to evaluate the effect of water flooding development. J Pet Explor Prod Technol 8, 1455–1463.
[45] Ma, S., Lv, M., Deng, F., Zhang, X., Zhai, H., Lv, W., 2015. Predicting the ecotoxicity of ionic liquids towards Vibrio fischeri using genetic function approximation and least squares support vector machine. J Hazard Mater 283, 591–598.
[46] Rodríguez-Pérez, R., Bajorath, J., 2020. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput-Aided Mol Des 34, 1013–1026.
[47] Chen, Y.-C., Tolbert, R., Aronov, A.M., McGaughey, G., Walters, W.P., Meireles, L., 2016. Prediction of Protein Pairs Sharing Common Active Ligands Using Protein Sequence, Structure, and Ligand Similarity. J Chem Inf Model 56, 1734–1745.
[48] Lu, Y., Meng, L., Ma, D., Cao, H., Liang, Y., Liu, H., Wang, Y., Jiang, G., 2021. The occurrence of PFAS in human placenta and their binding abilities to human serum albumin and organic anion transporter 4. Environ Pollut 273, 116460.
[49] Zhang, M., Rifas-Shiman Sheryl, L., Aris Izzuddin, M., Fleisch Abby, F., Lin Pi, I.D., Nichols Amy, R., Oken, E., Hivert, M.-F., 2023. Associations of Prenatal Per- and Polyfluoroalkyl Substance (PFAS) Exposures with Offspring Adiposity and Body Composition at 16–20 Years of Age: Project Viva. Environ Health Perspect 131, 127002.
[50] Park, S.K., Wang, X., Ding, N., Karvonen-Gutierrez, C.A., Calafat, A.M., Herman, W. H., Mukherjee, B., Harlow, S.D., 2022. Per- and polyfluoroalkyl substances and

incident diabetes in midlife women: the Study of Women's Health Across the Nation (SWAN). Diabetologia 65, 1157–1168.
[51] Ding, N., Karvonen-Gutierrez, C.A., Herman, W.H., Calafat, A.M., Mukherjee, B., Park, S.K., 2021. Associations of perfluoroalkyl and polyfluoroalkyl substances (PFAS) and PFAS mixtures with adipokines in midlife women. Int J Hyg Environ Health 235, 113777.
[52] Li, C.-H., Ren, X.-M., Cao, L.-Y., Qin, W.-P., Guo, L.-H., 2019. Investigation of binding and activity of perfluoroalkyl substances to the human peroxisome proliferator-activated receptor β/δ. Environ Sci: Process Impacts 21, 1908–1914.
[53] Sunderland, E.M., Hu, X.C., Dassuncao, C., Tokranov, A.K., Wagner, C.C., Allen, J. G., 2019. A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. J Expo Sci Environ Epidemiol 29, 131–147.
[54] Blake, B.E., Pinney, S.M., Hines, E.P., Fenton, S.E., Ferguson, K.K., 2018. Associations between longitudinal serum perfluoroalkyl substance (PFAS) levels and measures of thyroid hormone, kidney function, and body mass index in the Fernald Community Cohort. Environ Pollut 242, 894–904.
[55] Fedorenko, M., Alesio, J., Fedorenko, A., Slitt, A., Bothun, G.D., 2021. Dominant entropic binding of perfluoroalkyl substances (PFASs) to albumin protein revealed by ^{19}F NMR. Chemosphere 263, 128083.
[56] Wang, R., Lu, Y., Wang, S., 2003. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. J Med Chem 46, 2287–2303.
[57] Soni, L.K., Gupta, A.K., Kaskhedikar, S.G., 2009. Exploration of QSAR modelling techniques and their combination to rationalize the physicochemical characters of nitrophenyl derivatives towards aldose reductase inhibition. J Enzym Inhib Med Chem 24, 1002–1007.
[58] Benigni, R., Bossa, C., 2008. Predictivity of QSAR. J Chem Inf Model 48, 971–980.
[59] Chirico, N., Gramatica, P., 2011. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. J Chem Inf Model 51, 2320–2335.
[60] Tropsha, A., 2010. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol Inform 29, 476–488.
[61] Golbraikh, A., Tropsha, A., 2002. Beware of q2! J Mol Graph Model 20, 269–276.
[62] Devinyak, O.T., Slivka, M.V., Slivka, M.V., Vais, V.M., Lendel, V.G., 2012. Quantitative structure-activity relationship study and directed synthesis of Thieno [2,3-d]pyrimidine-2,4-diones as monocarboxylate transporter 1 inhibitors. Med Chem Res 21, 2263–2272.
[63] Liu, X., Yang, J., Guo, W., 2020. Semiempirical van der Waals method for two-dimensional materials with incorporated dielectric functions. Phys Rev B 101, 045428.
[64] Cruciani, G., Pastor, M., Guba, W., 2000. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci 11, S29–S39.
[65] Enache, M., Andriesei, B.M., Oancea, A., Udrea, A.-M., Raducan, A., Oancea, P., Avram, S., 2023. Interaction of anti-inflammatory drug nimesulide with ionic and non-ionic surfactant micelles: Insights from spectral and bioinformatics approach. J Mol Liq 392, 123511.
[66] Bjork, J.A., Wallace, K.B., 2009. Structure-activity relationships and human relevance for perfluoroalkyl acid–induced transcriptional activation of peroxisome proliferation in liver cell cultures. Toxicol Sci 111, 89–99.
[67] Datta, S., Dev, V.A., Eden, M.R., 2019. Using correlation based adaptive LASSO algorithm to develop QSPR of antitumour agents for DNA–drug binding prediction. Comput Chem Eng 122, 258–264.
[68] Akimoto, H., Uesawa, Y., Hishinuma, S., 2021. Molecular Determinants of the Kinetic Binding Properties of Antihistamines at the Histamine H1 Receptors. Int J Mol Sci 22, 2400.
[69] Palmer, D.S., Boyle, N.M.O., Glen, R.C., Mitchell, J.B.O., 2007. Random Forest Models To Predict Aqueous Solubility. J Chem Inf Model 47, 150–158.
[70] Kruger, F.A., Overington, J.P., 2012. Global Analysis of Small Molecule Binding to Related Protein Targets. PLOS Comput Biol 8, e1002333.
[71] Lee, D., Redfern, O., Orengo, C., 2007. Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8, 995–1005.
[72] Weiss, J.M., Andersson, P.L., Lamoree, M.H., Leonards, P.E., van Leeuwen, S.P., Hamers, T., 2009. Competitive binding of poly-and perfluorinated compounds to the thyroid hormone transport protein transthyretin. Toxicol Sci 109, 206–216.
[73] Zhang, L., Duan, X., Sun, W., Sun, H., 2020. Perfluorooctane sulfonate acute exposure stimulates insulin secretion via GPR40 pathway. Sci Total Environ 726, 138498.
[74] Hoyeck, M.P., Matteo, G., MacFarlane, E.M., Perera, I., Bruin, J.E., 2022. Persistent organic pollutants and β-cell toxicity: a comprehensive review. Am J Physiol-Endocrinol Metab 322, E383–E413.
[75] Thin, N., Hang, L., Thomas, P.Q., Tri, N., Thuc Duy, L., Svetha, V., 2021. GraphDTA: Predicting drug–target binding affinity with graph neural networks. Bioinformatics 37, 1140–1147.
[76] Chu, Z., Huang, F., Fu, H., Quan, Y., Zhou, X., Liu, S., Zhang, W., 2022. Hierarchical graph representation learning for the prediction of drug-target binding affinity. Inf Sci 613, 507–523.