

Article

How the Structure of Per- and Polyfluoroalkyl Substances (PFAS) Influences Their Binding Potency to the Peroxisome Proliferator-Activated and Thyroid Hormone Receptors—An In Silico Screening Study

Dominika Kowalska ^{1,†}, Anita Sosnowska ^{1,*,†}, Natalia Bulawska ¹, Maciej Stępnik ¹, Harrie Besselink ², Peter Behnisch ² and Tomasz Puzyn ^{1,3,*}

¹ QSAR Lab Ltd., Trzy Lipy 3, 80-172 Gdansk, Poland

² BioDetection Systems B.V., Science Park 406, 1098 XH Amsterdam, The Netherlands

³ Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

* Correspondence: a.sosnowska@qsarlab.com (A.S.); t.puzyn@qsarlab.com (T.P.)

† These authors contributed equally to this work.

Abstract: In this study, we investigated PFAS (per- and polyfluoroalkyl substances) binding potencies to nuclear hormone receptors (NHRs): peroxisome proliferator-activated receptors (PPARs) α , β , and γ and thyroid hormone receptors (TRs) α and β . We have simulated the docking scores of 43 perfluoroalkyl compounds and based on these data developed QSAR (Quantitative Structure-Activity Relationship) models for predicting the binding probability to five receptors. In the next step, we implemented the developed QSAR models for the screening approach of a large group of compounds (4464) from the NORMAN Database. The in silico analyses indicated that the probability of PFAS binding to the receptors depends on the chain length, the number of fluorine atoms, and the number of branches in the molecule. According to the findings, the considered PFAS group bind to the PPAR α , β , and γ only with low or moderate probability, while in the case of TR α and β it is similar except that those chemicals with longer chains show a moderately high probability of binding.

Keywords: perfluoroalkyl compounds; PFAS; peroxisome proliferator-activated receptor; thyroid receptor; QSAR; MLR; in silico; binding probability; molecular docking; virtual screening



Citation: Kowalska, D.; Sosnowska, A.; Bulawska, N.; Stępnik, M.; Besselink, H.; Behnisch, P.; Puzyn, T. How the Structure of Per- and Polyfluoroalkyl Substances (PFAS) Influences Their Binding Potency to the Peroxisome Proliferator-Activated and Thyroid Hormone Receptors—An In Silico Screening Study. *Molecules* **2023**, *28*, 479. <https://doi.org/10.3390/molecules28020479>

Academic Editors: Bono Lučić and Bakhtiyor Rasulev

Received: 8 December 2022

Revised: 22 December 2022

Accepted: 23 December 2022

Published: 4 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Per- and polyfluoroalkyl compounds (PFAS) are widely used in many consumer products and industrial applications, such as flame retardants [1], adhesives [2], varnishes [3], coatings [3], and food packaging materials [4]. PFAS are recognized by their hydrophobic carbon backbone, which is fully or partially saturated with fluorine atoms, and by possessing a hydrophilic functional group. PFAS, because of the C-F bonds, are very stable [2] and highly persistent in the environment, which can have a huge impact on humans and living organisms [5,6]. Continuous usage of PFAS leads to their accumulation in the environment and simultaneously increases the likelihood of harmful effects [2]. Many various PFAS have been detected so far in the air, water, soil, and dust [7–9]. Moreover, there are also studies indicating their presence in the human liver, blood, serum, and breast milk [10–13]. PFAS are known for causing different toxic effects including developmental, reproductive, carcinogenic, and immunological effects [14–16]. All these reports prompted scientists and regulatory bodies to take a closer look at the PFAS properties and behavior in the environment as well as the adverse effects they may pose. The results of the previous scientific research led to the introduction of the two most frequently used PFAS, i.e., PFOS and PFOA, into the Stockholm International Convention in 2009 and 2019, respectively [17,18]. All subsequent research and activities aimed at limiting/banning the use of PFAS compounds, but also at proposing new strategies to protect the environment and human health from

PFAS and PM (persistent and mobile) chemicals. These new strategies for monitoring the PFAS are the subject of the three European Horizon 2020 funded projects (PROMISCES (101036449), SCENARIOS (101036756), ZeroPM (101037509)).

Recently, studies indicated that some perfluoroalkyl substances may have been endocrine disrupting chemicals (EDCs) by interfering with hormone action [19,20]. These substances may act as agonists or antagonists of nuclear hormone receptors (NHRs). One of the groups of NHRs with which PFAS compounds may interact is peroxisome proliferator-activated receptors (PPAR) alpha, beta, and gamma [6,21]. They are members of the nuclear receptor superfamily which plays a significant role in the regulation of lipid, glucose metabolism, and cellular growth [22]. Three PPAR isoforms (α , β/δ , and γ) have a similar sequence but they are encoded by distinct genes and are expressed in different tissues. Depending on the isoform, PPARs have specific roles during development and in adult life [23]. PPAR α is mainly expressed in the liver, heart, intestine, and kidney [24]. PPAR α mediates its functions by affecting fatty acid transport, esterification, and oxidation [23,25]. PPAR β has a proven role in reverse cholesterol transport, wound healing, and cell proliferation and apoptosis [23]. It has been shown to be ubiquitously expressed in both humans and rodents and participates in fatty acid oxidation, but also has a role in the regulation of blood glucose levels [23,25]. PPAR γ is abundantly expressed in white adipose tissue, placenta, liver, heart, and skeletal muscles [26,27]. Several studies have shown activation of the PPAR isoforms by perfluoroalkyl acids in Cos-1 and 3T3-L1 cells using various binding and reporter constructs in cell-based assays [28,29].

The other, very important group of the NHR receptors that may be affected by EDCs are alpha and beta thyroid hormone receptors (TRs). Thyroid hormones are synthesized from tyrosine; hence, they are the amino acid derivatives [30]. The amino acids sequences of TR α and TR β are identical in 85%, and both show high homology across species [31]. Thyroid hormones play an important role in numerous physiological processes, such as bone remodeling, metabolism regulation, cardiac and neurophysiological functions [32–34]. PFAS may compete with thyroid hormones for the binding to the transporting proteins such as transthyretin or albumin, which may eventually increase the level of bioavailable hormone [35,36]. In people exposed to PFAS, both a decrease and an increase in thyroid hormones have been observed, and both are harmful to human health [37].

Recently, the development of computational methods has made it possible to virtually screen chemical compounds for various purposes [38,39]. Application of *machine learning techniques* can aid in assessing the potential toxicity of chemicals significantly reducing the time and cost associated with conducting biological experiments. One of the methods used in this study is molecular docking—a computational modeling technique that enables the prediction of the interactions between two molecules such as drugs, enzymes, or proteins [40]. Molecular docking can be successfully applied for the study of potential ligand (PFAS)-protein binding [41]. Second computational method widely used in virtual screening is QSAR (Quantitative Structure-Activity Relationship) modeling. QSAR modeling is based on the fact that the molecular structure of chemicals contains information about biological properties; hence, it can be used to predict such properties of chemicals without a need to perform experiments [42,43]. According to the QSAR models, it is possible to interpolate the activity of a group of chemicals, provided that the molecular descriptors were calculated for a whole dataset, and there is only part of a group with unknown activity [44,45].

The main purpose of this study is to indicate if there is a relationship between the chemical structure of PFAS (e.g., carbon chain length) and their binding to PPARs and TRs. Combining molecular docking with the QSAR model allows for studying the endocrine disruption with sufficient predictive power by analyzing the influence of PFAS structural features on docking scores. We have determined docking scores for 43 commonly used PFAS using the open-source Endocrine Disruptome tool [41]. Next, we have developed the QSAR models for predicting the binding scores of PFAS to receptors (PPAR α , PPAR β , PPAR γ , TR α , and TR β), and identified the structural features mainly affecting the binding

probability. In the further step, we have implemented the developed models for predicting the binding scores for 4464 different PFAS and thus for the virtual screening of influences of the PFAS structure on the binding probability to the NHRs. Due to very little experimental data on the nuclear receptor affinity of PFAS, the use of in silico studies is essential to learn more about the subject.

2. Results and Discussion

In the presented work we raised the important topic regarding the effects of the presence of PFAS in the environment and their impact on humans. We have applied an integrated computational approach, consisting of several steps allowing for virtual screening of PFAS in terms of their ability to bind to the NHRs. Firstly, using free-available software [41] we have calculated the binding scores of 43 PFAS to human nuclear hormone receptors (PPAR α , PPAR β , PPAR γ , TR α , and TR β). Then using obtained values, we developed and validated five QSAR models in accordance with the standards and recommendations of the Organization for Economic Co-operation and Development (OECD) [46–48]. Finally, we have performed a virtual screening of the binding potential of 4464 PFAS from the NORMAN Database to NHRs. The stages of our work are shown in Figure 1.

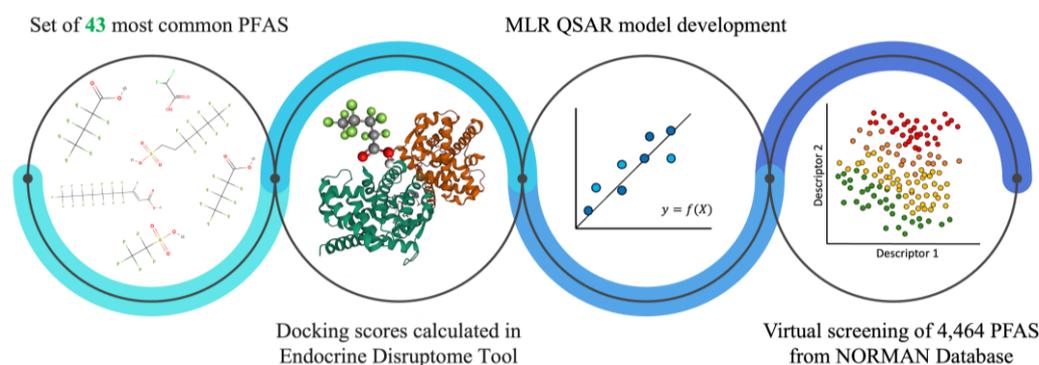


Figure 1. Workflow of research.

2.1. Predicting PFAS Binding Probability to PPAR (α , β , and γ) and TR (α and β) with QSAR

Based on PFAS binding probability predicted by Endocrine Disruptome Tool we have developed five independent QSAR models (for PPAR α , PPAR β , PPAR γ , TR α , and TR β). All developed models fulfill the OECD guidance recommendations dedicated to the principles for validation of the QSAR/QSPR models [46–48]. For the properly prepared QMRFs please refer to Supplementary Materials S2. Prepared QSAR models are scientifically valid and ready to be reproduced for the prediction of novel PFAS compounds [49]. In Table 1 we have summarized the equations and statistics obtained by developed models for the PFAS binding to the PPARs and TRs. All statistical parameters are described in Section 4.4 of this paper (Methodology: QSAR model calibration and validation).

Table 1. Statistical parameters of to PPAR α , PPAR β , PPAR γ , TR α , and TR β models.

Model	PPAR α	PPAR β	PPAR γ	TR α	TR β
Equation	PPAR α BS = -7.499 (± 0.067) – 0.947 (± 0.070) \times ICR – 0.394 (± 0.070) \times PW2	PPAR β BS = -8.248 (± 0.069) – 1.059 (± 0.071) \times ICR – 0.409 (± 0.071) \times X%	PPAR γ BS = -7.727 (± 0.052) – 1.099 (± 0.053) \times ICR – 0.398 (± 0.053) \times X%	TR α BS = -8.230 (± 0.070) – 0.454 (± 0.071) \times X% – 1.189 (± 0.071) \times ICR	TR β BS = -8.724 (± 0.054) – 0.137 (± 0.061) \times X% – 1.509 (± 0.061) \times TPC
n	29	29	33	33	29
k	14	14	10	10	14
R ²	0.917	0.924	0.949	0.924	0.970
RMSE _C	0.341	0.352	0.287	0.384	0.276
MAE _C	0.288	0.272	0.201	0.327	0.221

Table 1. Cont.

Model	PPAR α	PPAR β	PPAR γ	TR α	TR β
Q^2_{LOO}	0.878	0.903	0.940	0.908	0.955
RMSE_{CV}	0.350	0.398	0.310	0.422	0.340
MAE_{CV}	0.244	0.304	0.219	0.360	0.256
Q^2_{F1}	0.897	0.917	0.907	0.899	0.948
Q^2_{F2}	0.897	0.916	0.906	0.899	0.947
Q^2_{F3}	0.912	0.922	0.952	0.934	0.954
RMSE_{EXT}	0.352	0.359	0.279	0.359	0.344
MAE_{EXT}	0.257	0.279	0.223	0.312	0.280
CCC_{EXT}	0.937	0.952	0.953	0.946	0.973
Y^2_{SCR}	0.070	0.068	0.062	0.061	0.070

2.1.1. PPAR α , β , and γ

The best QSAR model for predicting the PFAS-PPAR α binding affinity utilized two molecular descriptors: radial centric information index (ICR) and path/walk 2-Randic shape index (PW2), between which the correlation coefficient was small ($r = 0.31$). ICR descriptor gives information about centrality in the molecules [50], whereas the PW2 is defined as (P2/W2), the quotient of the path length of 2 (P2) and walk length of 2 (W2) [51]. The results indicated a high correlation between calculated docking scores (using Endocrine Disruptome Tool) [41] and predicted by the developed QSAR model values (Figure 2A). The values of R^2 , Q^2_{LOO} , and $Q^2_{\text{F1,F2,F3}}$ are close to 1, which confirms that the predictions were accurate, the model is stable, and that the model has good predictive abilities. The model is characterized by relatively low values of the root mean square errors (Table 1) of prediction in the training and validation sets (respectively RMSE_{C} , RMSE_{CV} , and RMSE_{EXT}). More details about statistical results can be found in Supplementary Materials S1. To verify the reliability of predictions (they should be located within the optimum prediction space) we have checked the applicability domain of the developed model using the Williams plot (Figure 2B). This method allows to graphically present the standardized residuals (differences between observed and predicted values) versus the leverage value (indicates deviations of the structure of the compound from those used for the QSAR development). All in this work studied compounds were in the range of residuals differing by ± 3 standard deviations from the mean value ($h^* = 0.310$). One compound on the Williams plot, trifluoroacetic acid-TFA (43), has a higher leverage value than h^* , but its activity has been predicted correctly [52]. It is worth mentioning that TFA is one of ten carboxylic acids used to build and validate the model; however, it is the shortest one, which may be the reason of being for outlier. To prove that the model is not a “correlation-by-chance”, Y^2_{SCR} has been calculated (Table 1). Regarding the mechanistic interpretation of the model, PFAS with the lowest binding scores, 10:2 FTUCA (2) and 4:2 diPAPs (3) exceeding -8.9 kcal/mol, also have a high branching descriptor value (ICR) [53], which confirms that binding affinity is higher when it comes to long-chain perfluoroalkyl compounds (with more than eight carbon atoms in a molecule) [54].

We have estimated the binding probability of PFAS to PPAR β based on two molecular descriptors 2D ($r = 0.26$): radial centric information index (ICR) and the percentage of halogen atoms (X%). X% is a constitutional indices descriptor, which increases with the number of CF_2 groups in a molecule. The main reason for choosing a model based on this structure properties was the high similarity between calculated and predicted values (Figure 2C) and very good statistics (Table 1). The Williams plot (Figure 2D) indicates that all compounds were located within the optimum prediction space of the model; however, one point-TFA (43) has a leverage value higher than $h^* = 0.310$. However, attendance of this compound in the training set stabilizes the model. Nevertheless, the predictions for compounds with $h > h^*$ are treated as the results of extrapolation, so they are less reliable [55]. Scrambling test (Y^2_{SCR}) confirms that the presented model is statistically significant. Only five compounds (10:2FTUCA, 4:2diPAPs, PFDA, PFNS, and PFuDA) from the entire dataset show a moderate probability of binding to PPAR β (values are lower than

–9.6). Compounds with the highest binding scores are characterized by $X\%$ and ICR on a high level ($X\%$ not lower than 40.91, and ICR higher than 2.51).

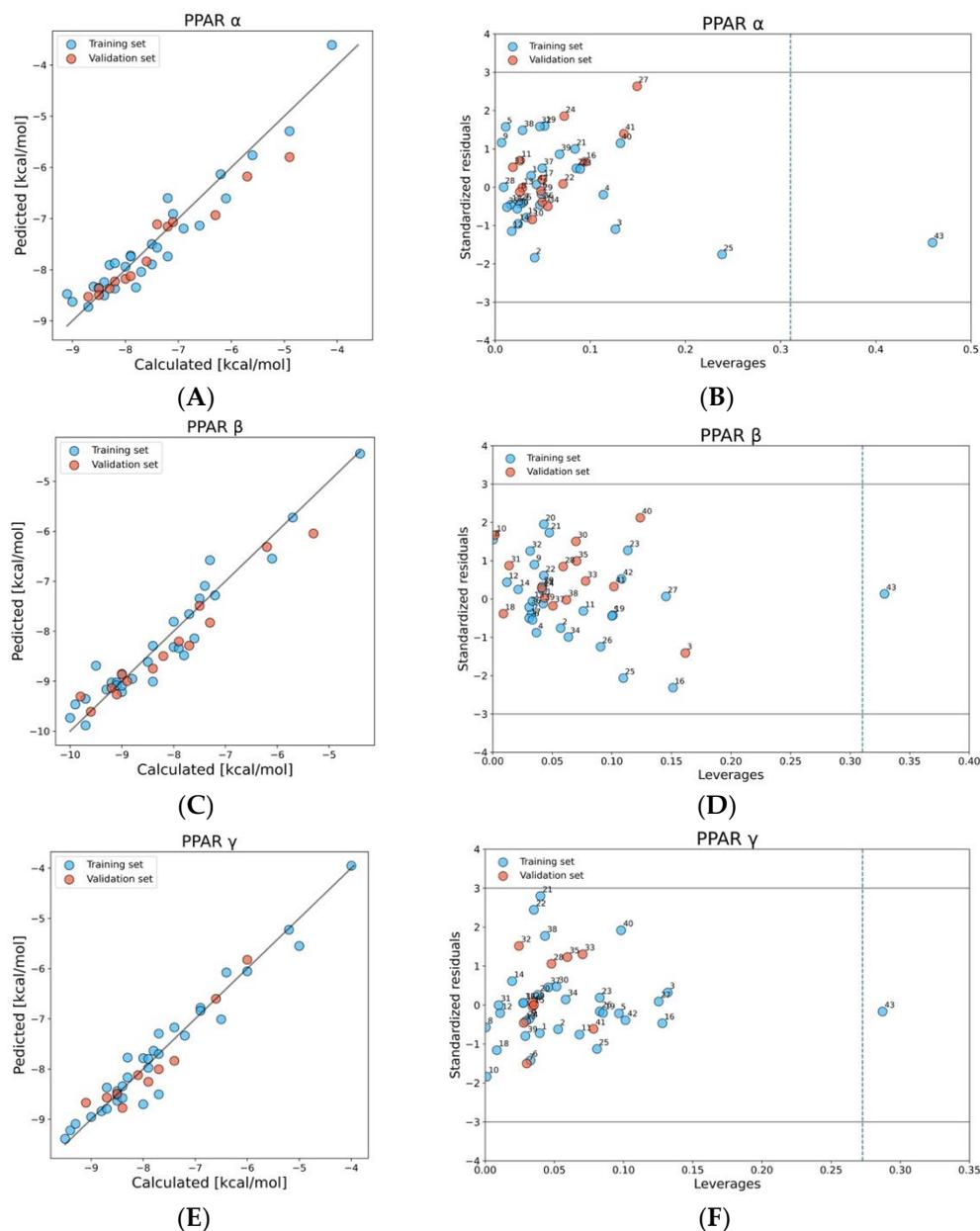


Figure 2. Plot of calculated versus predicted values of binding scores PPAR α , β , and γ (A,C,E). Williams plot: dash line indicates the critical leverage value, solid lines represent ± 3 standard deviation units (B,D,F).

To estimate PFAS binding probability to PPAR γ we have developed a QSAR model based on the same descriptors as for PPAR β ($X\%$ and ICR), with a correlation at $r = 0.26$. Figure 2E presents a significant correlation between the observed vs. predicted values of binding scores for PPAR γ . Model is characterized by satisfactory goodness-of-fit, robustness, and predictive capabilities (Table 1), which proves the accuracy of predictions. Values of errors (RMSE_C, RMSE_{CV}, RMSE_{EXT}) are also acceptable (Table 1). A small Y^2_{SCR} value confirms that the model is not a “correlation by chance”. Considering the applicability domain (Figure 2F), it can be seen that TFA (43) structurally diverges from the rest of the compounds ($h > h^* = 0.273$) but is situated in the range of residuals differing by ± 3 standard deviations from the mean value. The lowest binding score was observed

for PFUDa (42), which had the highest value of percentage of halogen atoms ($X\%$) and a relatively high value of ICR.

2.1.2. TR α and β

We developed the QSAR model for predicting PFAS-TR α binding probability based on two molecular descriptors: percentage of halogen atoms in a molecule ($X\%$) and radial centric information index (ICR). The correlation coefficient of the descriptors was low ($r = 0.16$). A good correlation between observed (calculated using Endocrine Disruptome Tool) [41] and predicted values can be seen on a scatter plot (Figure 3A). Additionally, the model is characterized by high accuracy of prediction and low values of the root mean square errors (Table 1). Compounds that were found between ± 3 standard deviations from the mean value on Williams plot and do not exceed the value of $h^* = 0.273$ are inside the structural space of the model (Figure 3B). In the case of TR α , only TFA (43) has been classified as an outlier due to the higher leverage value than h^* . To prove that the model is not a “correlation-by-chance”, Y^2_{SCR} has been calculated (Table 1). 8:2 FTUCA with a binding energy of -9.8 kcal/mol was identified as the compound with the highest probability of binding to the TR α ICR value of more than 2.5 and an $X\%$ of 53.33. Other compounds with a high probability of binding (binding score of less than -9.3 kcal/mol) had an $X\%$ value of no lower than 45 and an ICR of more than 2.24.

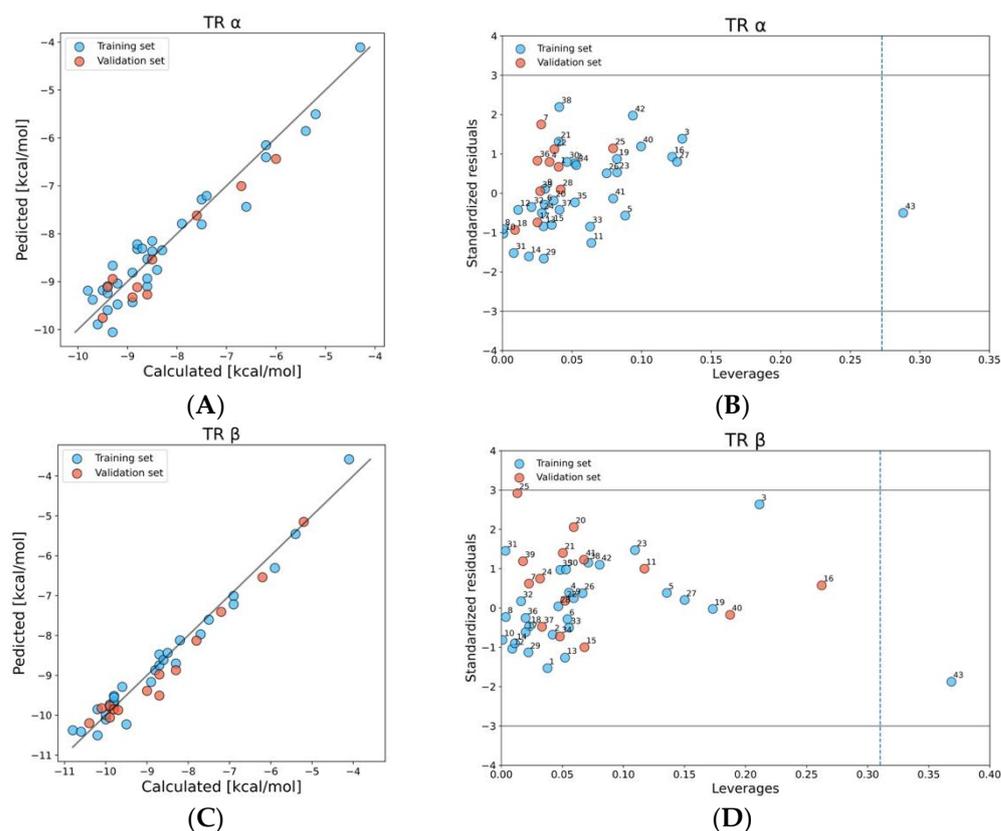


Figure 3. Plot of calculated versus predicted values of binding scores TR α , and β (A,C). Williams plot: dash line indicates the critical leverage value, solid lines represent ± 3 standard deviation units (B,D).

The best QSAR model for predicting PFAS binding probability to TR β is based on two relatively low correlated ($r = 0.47$) molecular descriptors: the percentage of halogen atoms in a molecule ($X\%$) and the total path count (TPC). TPC is the walk and path count descriptor, which describes the total number of paths, and reflects the size of the molecule as well as its complexity [56]. The Figure 3C confirms the good correlation between calculated and predicted data. We have selected the appropriate model based on robustness and

ability of prediction confirmed by R^2 , Q^2_{LOO} , and $Q^2_{\text{F1,F2,F3}}$ values which are close to 0.95 (Table 1). The analysis of the applicability domain using the Williams plot (Figure 3D) shows that the same chemical as in the case of TR α -TFA (43) has a higher average hat value than $h^* = 0.273$; however, it does not have standard residuals greater than 3 standard deviation units ($\pm 3\sigma$). Thus, it can be stated that TFA (43) is slightly structurally different from the rest in the set, and has a positive effect on the extension of the model applicability domain. In the training set, two compounds, 10:2 FTOH (1) and 10:2 FTUCA (2), had the highest probability of binding to TR β (binding score lower than -10.6 kcal/mol) with having one of the highest TPC values in the set (6.39), and an X% value exceeding 50.

2.2. Screening of the Binding Probability of 4464 PFAS

After developing the models, we implemented them into an external dataset from the NORMAN Database System [57] to predict the binding probability based on selected structural features of PFAS. We have estimated the endocrine action against five receptors for 4464 PFAS and have assessed if the predictions are in the applicability domain (please refer to Supplementary Materials S1). The number of compounds that fall within the applicability domain of developed models was verified using leverage value (it should be lower than h^*). Compounds were also outliers when they did not fall within the range of values predicted by the model. Next, we analyzed the relationships between the particular receptor binding strength and the chemical structure of PFAS. Additionally, we have selected 59 PFAS which were from five different structural groups (carboxylic acids-group C, sulfonic acids-group S, phosphonic acids-group P, fluorotelomer alcohols-group F, and dicarboxylic acids-group Y) to illustrate how the structure of PFAS influences the value of the docking score between PFAS and receptors. Within the groups, each subsequent chemical differs from the previous one by an additional CF_2 group. The increasing number of CF_2 groups affects the change in the values of descriptors, while also influencing binding scores.

2.2.1. PPAR α , β , and γ

With all predictions for PFAS selected for the virtual screening, 1799, 2234, and 2631 compounds were outside the applicability domain of the developed models for PPAR α , PPAR β , and PPAR γ , respectively. Information about all compounds outside the structural AD is attached in Supplementary Materials S1. Based on docking scores obtained in models, we have assigned each compound to classes by color: “red” indicates a high probability of binding, “orange” indicates a moderately high binding probability, “yellow” corresponds to a moderate probability of binding, and “green” corresponds to a low probability of binding (more details in Supplementary Materials S1). To visualize the relationship between the chemical structure of the PFAS and their binding potency to PPAR receptors we have prepared the scatter plots, where the x and y axes represent the descriptors used in the QSAR models (Figure 4, Figure 5 and Figure 7).

Analyzing the results obtained for PPAR α (Figure 4A), all predictions for PFAS falling into the applicability domain (2665 chemicals) created only one class of low binding probability (green) to PPAR α (more details are in Supplementary Materials S1). The docking scores for compounds within AD estimated for this endpoint range from -8.78 to -4.49 , and the $h^* = 0.310$. For compounds with higher values of PW2 (close to 0.7) and ICR (close to 3), the docking scores decrease and have values of about -8.7 . PW2 is a descriptor that gives information related to the shape of compounds and ICR informs about two-dimensional centrality in the molecules. The results obtained during the analysis suggest that all compounds similar to the training set of the QSAR model (based on PW2 and ICR) will have a low binding probability to PPAR α . The above results indicate that PFAS are unlikely to bind to the PPAR α receptor [22]. However, the compounds in the NORMAN database include many specific elements and groups (e.g., Ag, Ba, Cd, Ce, Cs, Si, I, Br, etc.), so it cannot be stated unequivocally that these compounds will not bind to PPAR α .

In the case of the PPAR α , we can state that increasing the length of the carbon chain increases the PW2 and ICR values (Figure 4A). Within all groups (C, S, P, F, and Y), a tendency can be seen for the probability of binding to increase with increasing descriptors values (Figure 4C). However, all compounds remain in the class with a low probability of binding to the receptor (green), because of the similarity of compounds to the training set (Figure 4B). The conclusion is consistent with the equation of the model, where the regression coefficient of PW2 and ICR is negative.

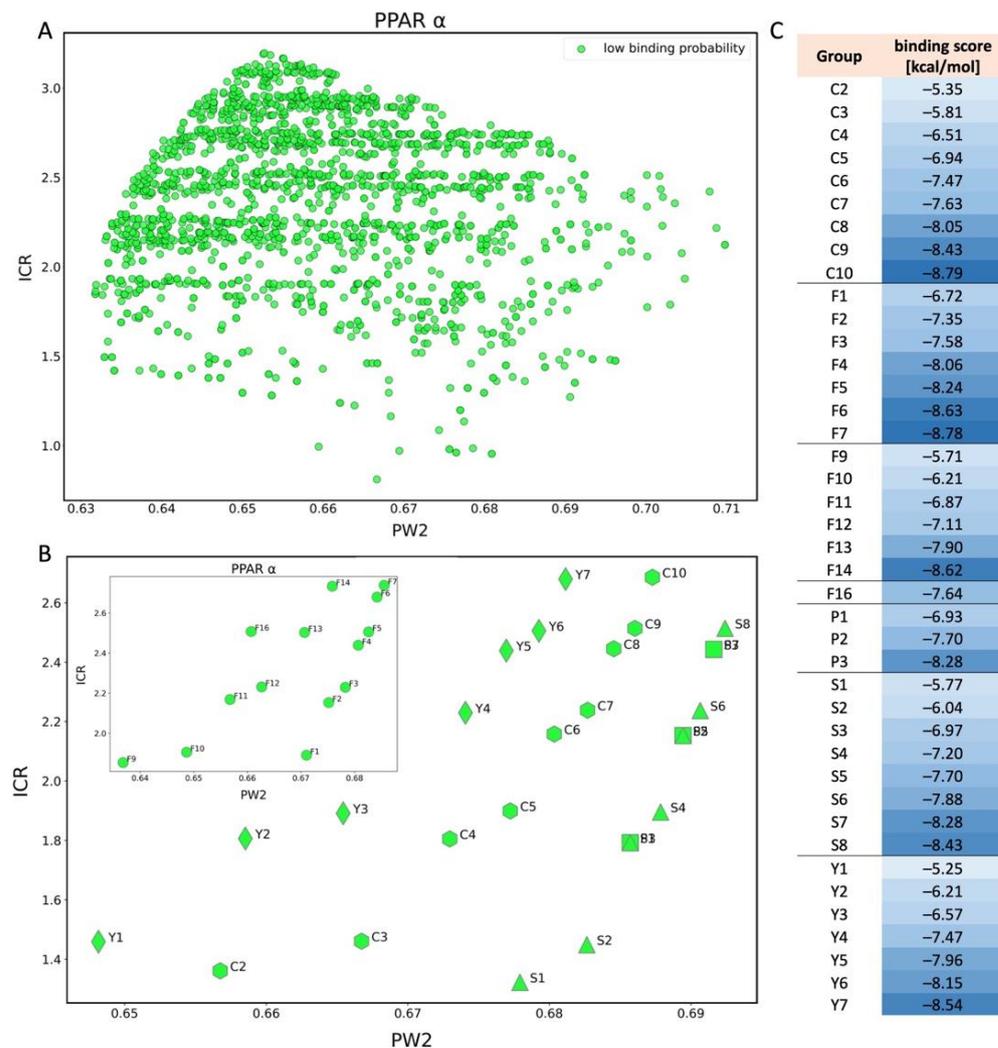


Figure 4. (A) Scatter plot of binding scores obtained for PPAR α in PW2 and ICR space; (B) scatter plot for PPAR α showing the division of PFAS into five different structural groups in space of PW2 and ICR. C-carboxylic acids (COOH-(CF₂)_n-CF₃), S-sulfonic acids (SO₃H-(CF₂)_n-CF₃), P-phosphonic acids (PO₃H₂-(CF₂)_n-CF₃), F-Fluorotelomer alcohol (F1-F7(OH-(CH₂)-(CF₂)_n-CF₃), F8-F14 (OH-(CH₂)₂-(CF₂)_n-CF₃), F16 (OH-(CH₂)₃-(CF₂)_n-CF₃), F17 (OH-(CH₂)₄-(CF₂)_n-CF₃)), Y-dicarboxylic acids (COOH-(CF₂)_n-COOH). Colors: green-low binding probability; (C) Table showing binding score values for each compound shown in chart B.

Several experimental studies can be found in the literature describing the effect of PFAS on PPAR α , however, these studies vary in terms of the technique used to perform the tests, the type of cells tested, and the conditions under which the experiments were conducted. For example, Takacs and Abbott [23] who used transiently transfected Cos-1 cells (from African green monkey kidney) to evaluate the potential of PFOA and PFOS to activate human PPAR α , found that PFOA was able to significantly increase the PPAR α activity. The same authors reported also in vitro studies [24], which show that, e.g., perfluorodecanoic

acid-PFDA and perfluorononanoic acid-PFNA bind more strongly to the PPAR α than perfluorohexanoic acid-PFHxA, indicating that response depends on chain length and functional group. It is in line with our studies, where PFDA and PFNA have a lower value of docking scores (-8.5 and -8.0 , respectively) than PFHxA (-7.1), which indicates that even if the PFDA and PFNA remain in a group with a low binding probability to PPAR α , they bind strongly than PFHxA.

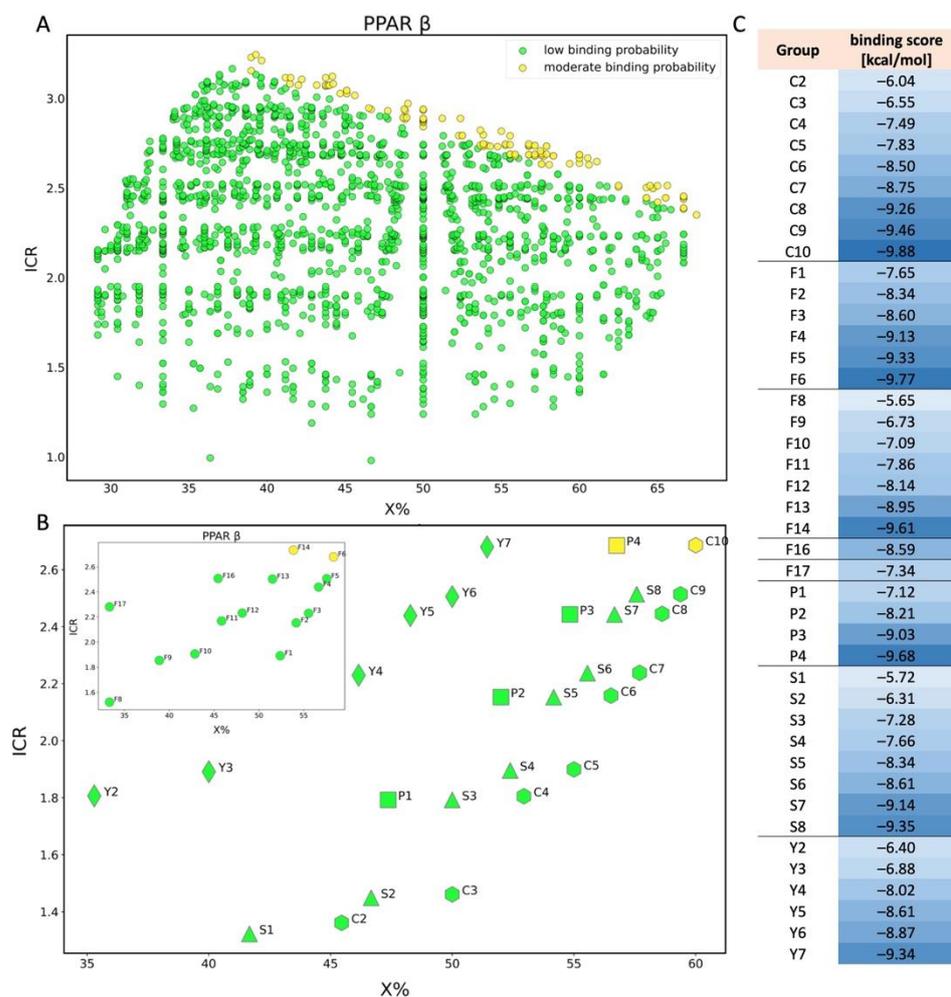


Figure 5. Scatter plot of binding scores obtained for PPAR β in X% and ICR space (A); Scatter plot for PPAR β showing the division of PFAS into 5 different structural groups in space of X% and ICR. C-carboxylic acids (COOH-(CF₂)_n-CF₃), S-sulfonic acids (SO₃H-(CF₂)_n-CF₃), P-phosphonic acids (PO₃H₂-(CF₂)_n-CF₃), F-Fluorotelomer alcohol (F1-F7(OH-(CH₂)-(CF₂)_n-CF₃), F8-F14 (OH-(CH₂)₂-(CF₂)_n-CF₃), F16 (OH-(CH₂)₃-(CF₂)_n-CF₃), F17 (OH-(CH₂)₄-(CF₂)_n-CF₃)), Y-dicarboxylic acids (COOH-(CF₂)_n-COOH). Colors: green-low binding probability, yellow-moderate binding probability (B); table showing binding score values for each compound shown in chart B (C).

Regarding the screening for PPAR β , we have classified the predictions into two groups: low binding probability (green-2108 PFAS) and moderate binding probability (yellow-122 PFAS). In Figure 5A we have presented the correlation between ICR (corresponds to a number of branching), X% (percentage of halogen atoms), and PFAS binding potency to PPAR β . The docking scores estimated for this endpoint range from -9.88 to -9.60 for the moderate binding probability class (yellow) and from -9.59 to -4.67 for the low binding probability class (green). An analysis of the descriptor values shows that compounds with more halogen atoms (high X% value), and more branches in the main chain have lower binding scores (they bind more strongly to the receptor) [58]. Worth considering is that QL01313 (Figure 6), despite the relatively low content of halogen atoms (X% = 40%), was in

the group of compounds with a moderate probability of binding due to the high value of branching (ICR = 3.24). However, in the same group, there is also less branched compound (ICR = 2.35): QL03507 (Figure 6), which consists of 12 atoms of carbon and contains a lot of fluorine atoms in its structure (X% = 67.5%). Both mentioned PFAS were classified into a group with a moderate probability of binding (Figure 5A).

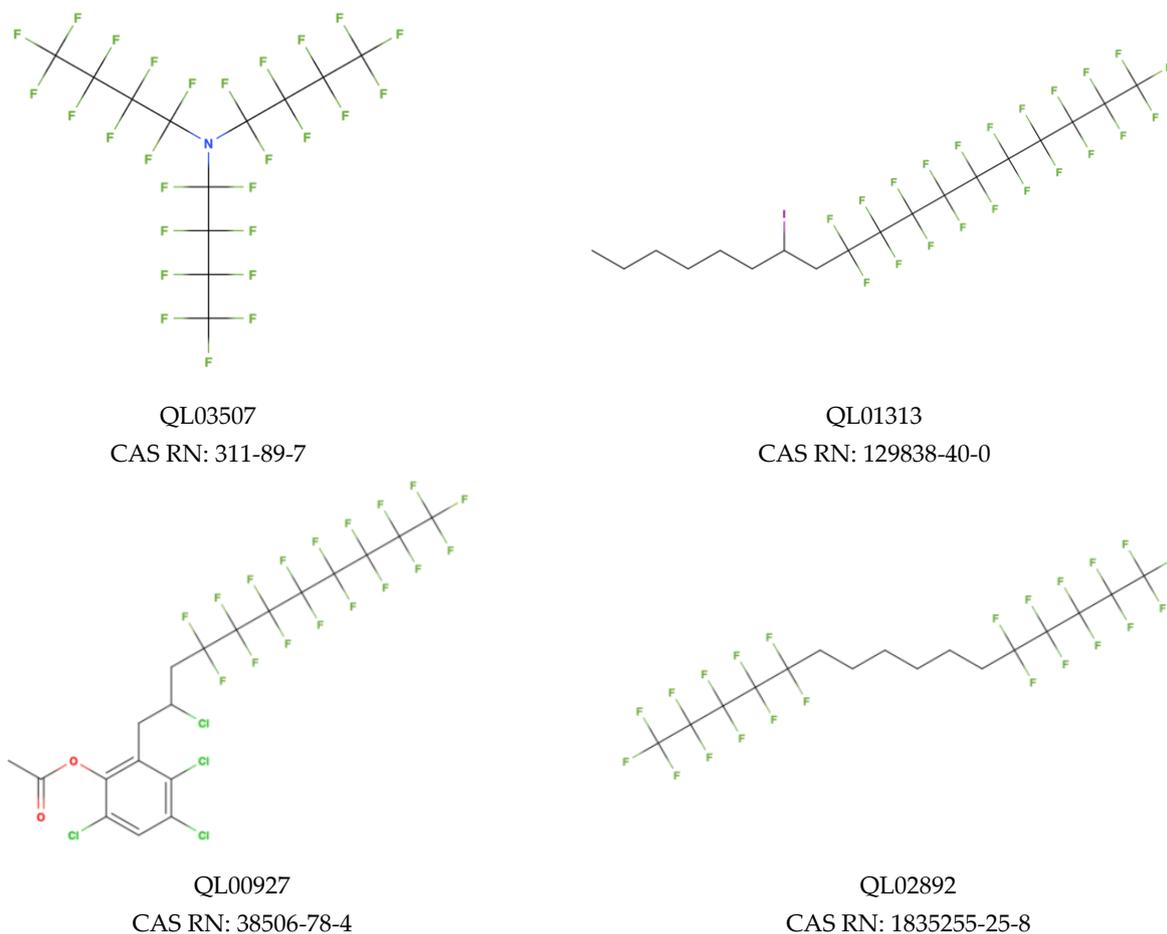


Figure 6. Chemical structure of PFAS selected from NORMAN Database.

In the case of PPAR β , for all chemical groups (perfluorinated carboxylic acids, fluoroalcohol alcohols, phosphonic acids, and dicarboxylic acids), an increase in binding probability is observed as the molecule is extended by another $-\text{CF}_2$ group (Figure 5C). Perfluoroalkyl carboxylic acid (C10) and phosphonic acid (P4) with X% close to 60% and ICR higher than 2.6 belong to the class of compounds with a moderate probability of binding (yellow). It can be seen in Figure 5B that long-chain PFAS bind stronger to PPAR β than to PPAR α , because the same compounds (C10, P4, F6, and F14) that belong to the group with moderate binding probability to PPAR β (yellow) have a lower binding probability to PPAR α (green). Two perfluoroalcohol alcohols (F14 and F6) differ by one $-\text{CH}_3$ moiety and are classified as PFAS with a moderate binding probability to PPAR β because both consist of 10 carbon atoms fully saturated with fluorine. Furthermore, it can be concluded that all considered perfluoroalkyl dicarboxylic acids (Y, which contain less than 10 carbon atoms saturated with fluorine) are classified as PFAS with low binding probability to PPAR β . None of the analyzed perfluoroalkyl sulfonic acids shows a significant probability of binding.

The impact of PFAS on the beta isoform of PPAR has also already been considered in the literature. Takacs and Abbott [23] in their study on transiently transfected Cos-1 cells indicated that neither PFOA nor PFOS affected PPAR β . Moreover, Heuvel et al. [29] in 3T3-L1 mouse fibroblasts assay found that both PFOA and PFOS could not activate

human PPAR β . Contradictory conclusions were reached by Li et al. [58] who investigated that the binding potency and transcriptional activity of PFOA, PFOS, and other PFAS toward human PPAR β (PDB: 3OZ0) increase with increasing carbon chain length. The authors found that differences in the cell type and reporter system may lead to different conclusions. Results obtained in the mentioned studies are in line with our results, where 10:2FTUCA, 4:2diPAPs, PFDA, PFNS, and PFUDa belong to the group with moderate binding probability, and the other PFAS with shorter chains have a low binding probability of binding to PPAR β .

Results obtained for the PPAR γ receptor indicate that the predictions of binding are classified into two groups: low binding probability (green-1633 PFAS) and moderate binding probability (yellow-200 PFAS). Figure 7A presents the scatter plot in the space of ICR (radial centric information index) and X% (percentage of halogen atoms). As in previous cases, assignment to groups is based on molecular descriptors values (X% and ICR), and when they are large it indicates a long-chain molecule, which can determine a higher probability of binding to the receptor [24]. More compounds from NORMAN Database show moderate binding probability (are in the yellow group) compared to PPAR β . For example, QL00927 and QL02892 (Figure 6) belong to the group of compounds with moderate binding probability to PPAR γ , while there is a low probability of binding to PPAR β for the same PFAS (for more details please refer to Supplementary Materials S1). It is due to the differences between the threshold values set in the tool based on statistical analysis for PPAR β and PPAR γ (Table 1).

In the case of the PPAR γ , we can state that increasing length of the carbon chain increases the PW2 and ICR values (Figure 7B). Considering the different functional groups of PFAS (C, S, P, Y, F), the tendency can be seen that as the value of descriptors increases, the probability of binding to PPAR γ increases (Figure 7C). Both perfluoroalkyl carboxylic acid (C10) and phosphonic acid (P4) belong to the class of compounds with a moderate probability of binding (yellow) to PPAR β . Interestingly, the C9 compound with the lower ICR value (2.5) shows a moderate binding probability, indicating that it is more likely to bind to the gamma isoform rather than to beta. Two perfluorotelomer alcohols (F14 and F6) are classified as PFAS with a moderate binding probability to PPAR γ (F14 has one -CH₃ group more, but both consist of ten carbon atoms fully saturated with fluorine). Despite dicarboxylic acid Y2 containing three carbon atoms fully saturated with fluorine and Y7 containing nine, all perfluoroalkyl dicarboxylic acids (Y2-Y7) are classified as PFAS with low binding probability to PPAR γ , for which their binding score decreases when number of C-F bonds in the main chain increases (Figure 7C). None of the analyzed perfluoroalkyl sulfonic acids shows a significant probability of binding to PPAR γ isoform.

Researchers have also paid attention to the dysfunction of PPAR γ after binding to PFAS. Zhang et al. [59] reported that 11 perfluorinated carboxylic acids and three perfluorinated sulfonic acids were able to activate human PPAR γ , which activity increased with increasing carbon chain length of PFAS. However, results obtained by Behr et al. [22] indicate that PFBA, PFHxA, PFBS, PFOA, and PFOS do not affect PPAR γ . Our study confirms the effect of chemical chain length on the probability of binding to the receptor; however, our analysis also shows that the compounds studied by Behr et al. have a low probability of binding to the PPAR γ (green). The differences probably result from using a different cell line and the transfection procedure in the case of Zhang's studies. Zhang et al. also specified that molecular docking has been performed on the specific 3D structure of PPAR γ (PDB entry:3U9Q), while Behr et al. did not include that information. In the case of our study, PDB entry:3ET3 has been taken into consideration, which also affects the differences in results.

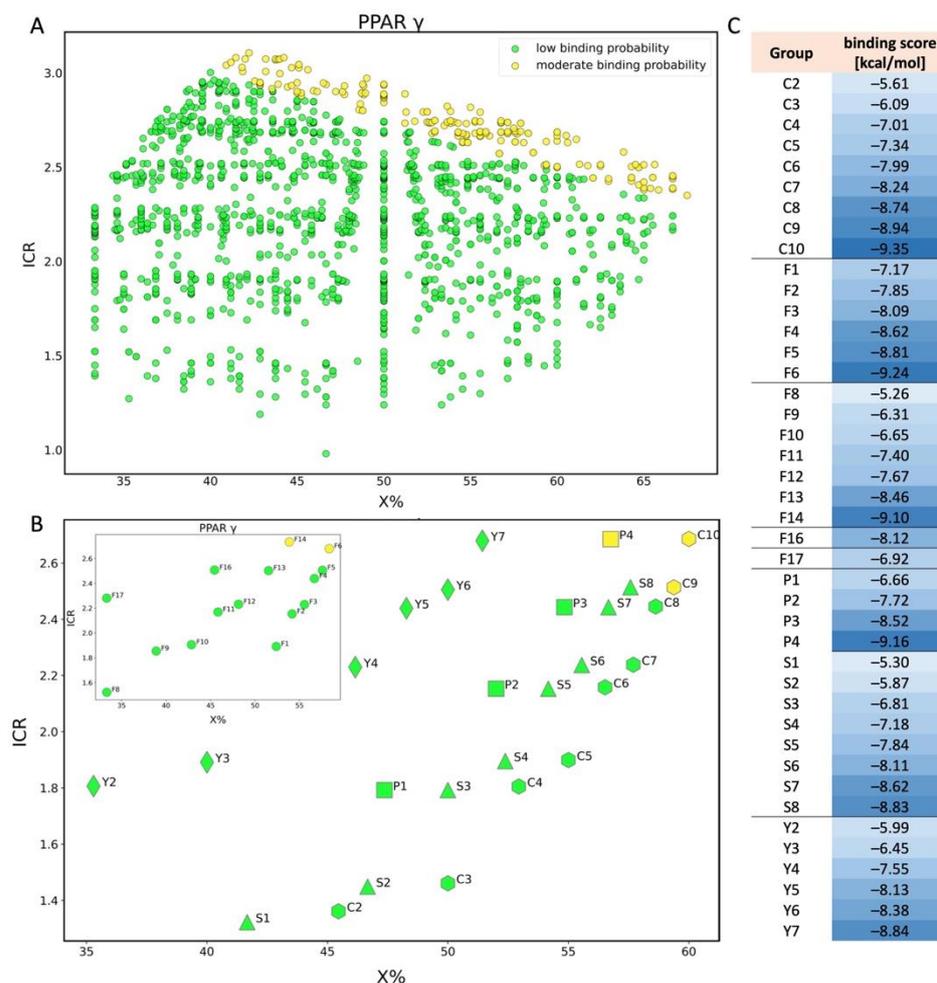


Figure 7. Scatter plot of binding scores obtained for PPAR γ in X% and ICR space (A); scatter plot for PPAR γ showing the division of PFAS into 5 different structural groups in space of X% and ICR. C-carboxylic acids (COOH-(CF₂)_n-CF₃), S-sulfonic acids (SO₃H-(CF₂)_n-CF₃), P-phosphonic acids (PO₃H₂-(CF₂)_n-CF₃), F-Fluorotelomer alcohol (F1-F7 (OH-(CH₂)-(CF₂)_n-CF₃), F8-F14 (OH-(CH₂)₂-(CF₂)_n-CF₃), F16 (OH-(CH₂)₃-(CF₂)_n-CF₃), F17 (OH-(CH₂)₄-(CF₂)_n-CF₃)), Y-dicarboxylic acids (COOH-(CF₂)_n-COOH). Colors: green-low binding probability, yellow-moderate binding probability (B); table showing binding score values for each compound shown in chart B (C).

2.2.2. TR α and β

For TR α , the minimum and maximum values of the estimated endpoint were -10.06 and -4.36 kcal/mol respectively, and the hat value obtained was $h^* = 0.273$. Compounds that were found between the minimum and maximum value of predicted binding energy for the training set and do not exceed the value of h^* are classified as a compound in the structural space of the model. We have classified 2104 compounds as outliers (52.9% coverage). Information about all compounds outside the structural AD is attached in Supplementary Materials S1. Compounds falling into the applicability domain of the TR α model (Figure 8A) represent three classes of binding: low binding probability (504 chemicals), moderate binding probability (1400 chemicals), and moderately high binding probability (456 chemicals). Based on standardized coefficients used in the QSAR model we can indicate that ICR is a descriptor that has a greater influence on the outcome of the estimated value of binding probability to the receptor than X%. In the first class (green-low binding probability) the range of values for the X% is from 29.03 to 62.5, and for the ICR it is from 0.92 to 2.29. Compound QL03536 (Figure 9), with the highest value of energy (-4.36 kcal/mol), has a low value of ICR (0.99) and X% value that equals 36.36, and is a cyclic structure with four halogen atoms in the structure. On the other hand, the lowest

value of energy (-7.19 kcal/mol) has a molecule QL00917 (Figure 9) with a high ICR value (2.27) resulting from a large number of branching, and the X% equals only 32.69. In the second class (yellow–moderate binding probability) the range of X% values are from 29.73 to 67.74, and of ICR is from 1.52 to 3.04. Compound QL03467 (Figure 9) is the carboxylic acid with almost all carbon atoms fully substituted by fluorine, and it has the highest value of energy (-7.20 kcal/mol), one of the lowest values of ICR (1.79), and X% equals 50.00. The lowest value of energy (-9.19 kcal/mol) has molecule QL01468, with a high value of ICR due to the main chain of 15 atoms and X% equals 37.78. In the class with moderately high binding probability, the compound with the lowest energy value (-10.06 kcal/mol) is QL00701 (Figure 9). It has a relatively high value of X% (57.89) because almost all atoms are completely substituted with fluorine and the value of ICR is equal to 2.75 which is close to the mean value of this descriptor in the group.

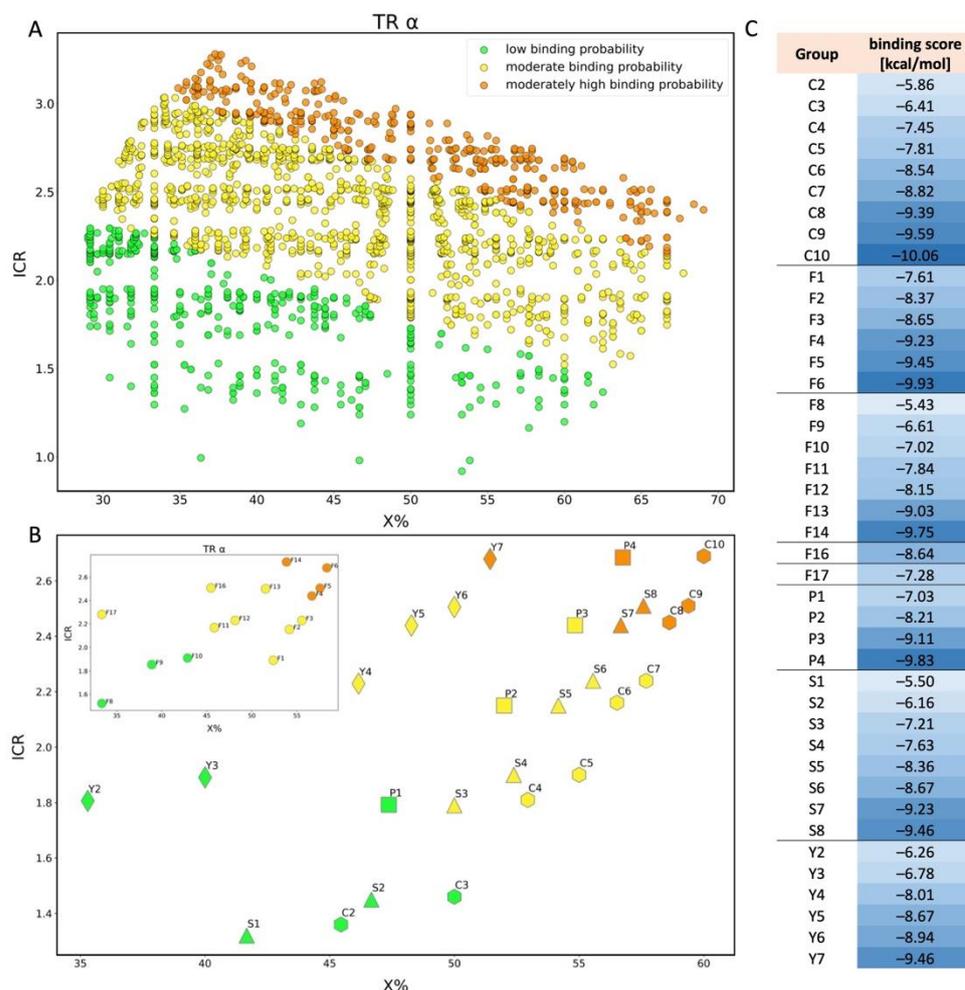


Figure 8. Scatter plot of binding scores obtained for TR α in X% and ICR space (A); scatter plot for TR α showing the division of PFAS into 5 different structural groups in space of ICR and X%. C-carboxylic acids (COOH-(CF₂)_n-CF₃), S-sulfonic acids (SO₃H-(CF₂)_n-CF₃), P-phosphonic acids (PO₃H₂-(CF₂)_n-CF₃), F-Fluorotelomer alcohol (F1-F7(OH-(CH₂)-(CF₂)_n-CF₃), F8-F14 (OH-(CH₂)₂-(CF₂)_n-CF₃), F16 (OH-(CH₂)₃-(CF₂)_n-CF₃), F17 (OH-(CH₂)₄-(CF₂)_n-CF₃)), Y-dicarboxylic acids (COOH-(CF₂)_n-COOH). Colors: green-low binding probability, yellow-moderate binding probability, orange-moderately high binding probability (B); table showing binding score values for each compound shown in chart B (C).

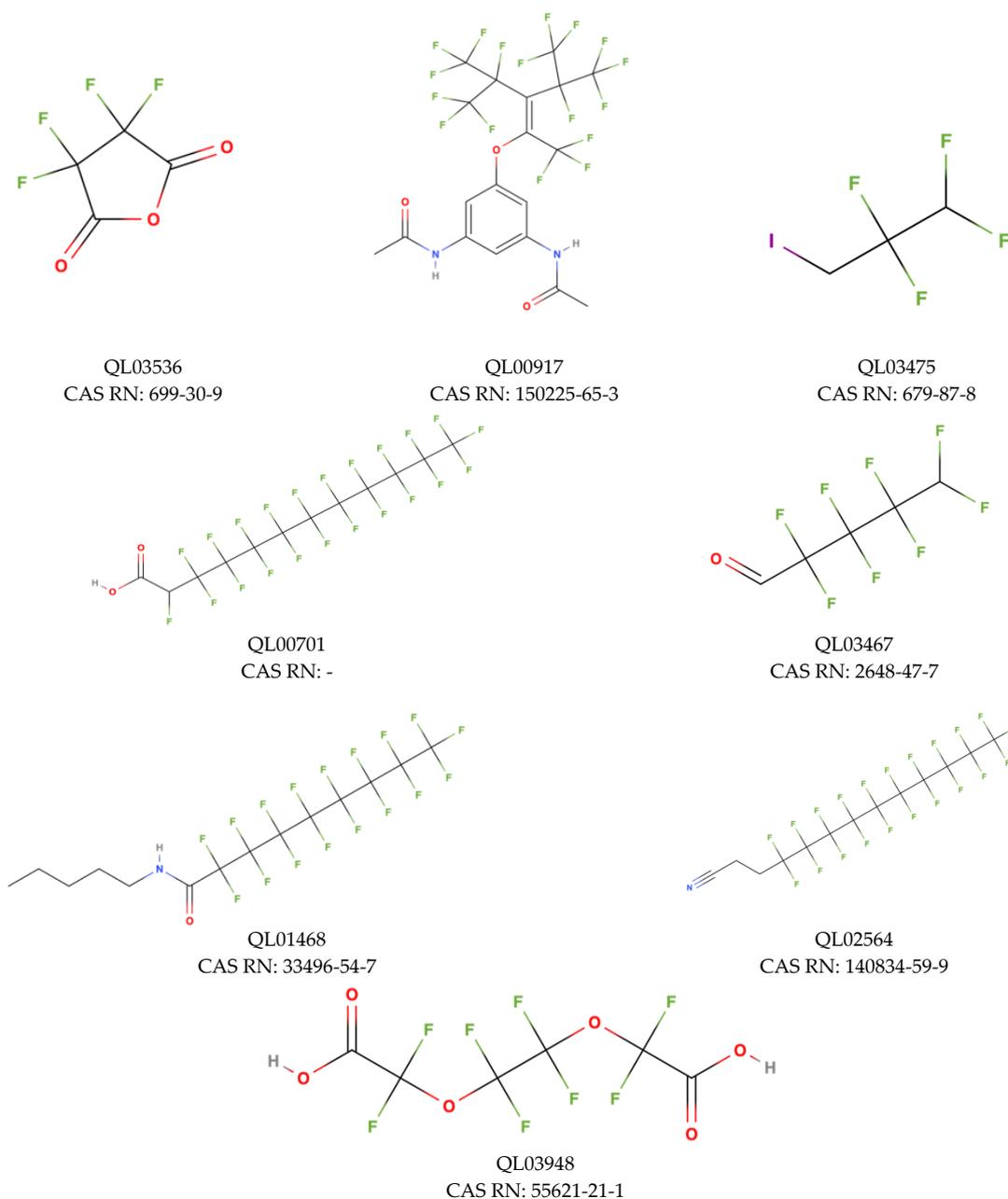


Figure 9. Chemical structure of PFAS selected from NORMAN Database.

Analyzing PFAS from different groups (Figure 8B) we can observe that carboxylic acid with five carbon atoms (C4) has a moderate binding probability, whereas acids composed of nine or more carbon atoms (C8-C10) are classified as a moderately high binding probability. In the case of the Y group, dicarboxylic acid with six carbon atoms in structure (Y3) is classified as a compound with a low probability of binding. Y4, which is first in the moderate binding probability class, has eight carbon atoms in the structure, while acid with moderately high binding probability (Y7) has 11 carbon atoms in the structure. All presented above examples confirm the higher probability of binding to a receptor in the case of long PFAS than for shorter compounds. Based on Figure 8C, it can be observed that extending the inner chain (+CF₂) of groups causes a decrease in the value of the binding score, although the compounds are sometimes assigned to the same class.

Many studies have focused on the detrimental effects of PFAS on the thyroid system [60], but few have looked at the docking of these compounds to thyroid α and β receptors. Our study consisted of examining the probability of PFAS binding to thyroid

α and β sites, respectively: 3ILZ (structure of TR α bound to selective thyromimetic GC-1 in P212121 space group [61]) and 3IMY (structure of TR β bound to selective thyromimetic GC-1 [62]). The binding of a ligand (GC-1) to sites selected for PFAS docking induces positive effects, such as lowering serum LDL cholesterol and triglycerides, so PFAS binding may interfere with this process [63].

The *in vitro* studies reported in the literature prove that short-chain PFAS such as perfluoropentanoic acid–PFPeA, perfluorobutanoic acid–PFBA, and perfluorobutanesulfonic acid–PFBS have a lesser effect on TR dysfunction (reduced cell viability of rat-thyroid cells [64]) compared to long-chain (more than eight carbon atoms) PFAS. In our set, the energy of binding for those three studied short-chain PFAS does not exceed -7.4 kcal/mol, which confirms the above statement. Another *in vitro* study documented the accumulation of PFOA and PFOS in thyroid cells [65], which, according to our study, have low binding energy values (-8.9 and -9.4 kcal/mol, respectively), meaning they are more likely to bind to the receptor. These long-chain compounds were considered less safe for human health and the environment.

Analyzing the screening approach for the TR β , the minimum and maximum values of the estimated endpoint were -10.5 and -4.27 kcal/mol respectively, and the hit value obtained was $h^* = 0.310$. The 2694 compounds were classified as outliers (39.7% coverage). Considering the results obtained for TR β , similarly to TR α , we have classified compounds into three classes (Figure 10A): low binding probability (477 chemicals), moderate binding probability (731 chemicals), and moderately high binding probability (562 chemicals). Here also the descriptor based on the size of the molecule (TPC) has a greater influence on the binding energy value, regarding the QSAR model. The low binding probability class consists of compounds with a range from 31.25 to 62.50 for X% and a range from 3.61 to 5.36 for TPC. A molecule QL03475 (Figure 9), with the highest value of energy (-4.27 kcal/mol) has a TPC equal to 3.61, and a quite high value of X% (45.45). In contrast, a molecule QL03948 (Figure 9) with the lowest value of energy (-7.80 kcal/mol) has TPC equal to 5.35 due to the eight atoms in the main chain and even lower X% (36.36) because only eight fluorine atoms are present in the molecule. The group of compounds with moderate binding probability includes molecules with X% in the range from 32.14 to 66.67 and TPC in the range from 5.15 to 6.09. In the third class (moderately high binding probability) compound QL02564 (Figure 9) has the lowest value of binding energy (-10.5 kcal/mol) with a TPC value equal to 6.45 due to the 13 atoms in the main chain and a relatively high value of X% (53.85) because of 10 fully fluorinated carbons.

In the case of TR β intra-group classification (Figure 10B), sulfonic acid with five carbon atoms in the structure (S4) is classified as a group with moderate binding probability, and eight or more carbon atoms in a molecule (S7-8) it is classified as a moderately high binding probability. For all other groups (perfluorinated carboxylic acids, fluorotelomer alcohols, phosphonic acids, and dicarboxylic acids), a decrease in the value of the binding score parameter can also be observed as the molecule is extended by another $-\text{CF}_2$ group (Figure 10C).

For the TR β receptor, there are no studies available that describe the binding probability of PFAS to the docking site we studied (3IMY). Young et al. [66] in their work presented a classification of selected PFAS as active (PFOS, FOSA, PFNA, PFDA) and as inactive (PFHxA, PFOA, PFHpA, and PFUnDA) antagonists against thyroid hormone receptor β . Our docking results classified PFOS, FOSA, PFNA, and PFDA compounds to second class with moderately high binding probability (binding score is not higher than -9.6 kcal/mol). Behnisch et al. [67] have presented research based on the TTR-TR β -CALUX[®] assay in which the activity of 13 PFAS was demonstrated, which indicates a potential for disturbing the functioning of thyroid hormones. Potency factors were expressed as PC₈₀-based and IC₅₀-based using PFOA as a reference compound. Based on the PC₈₀ values, short-chain compounds such as PFBS or PFBA and the longest compound tested (PFDCa) showed lower activity than the reference compound. Higher activity than PFOA was reported for PFHpA, PFHxS, and PFOS. It is not possible to directly compare the docking results we

obtained with the study conducted in the above paper, due to the completely different nature of the experiment. However, based on the binding score values we calculated for the PFAS group studied here, PFOS can be identified as the compound with the highest probability of binding to the receptor.

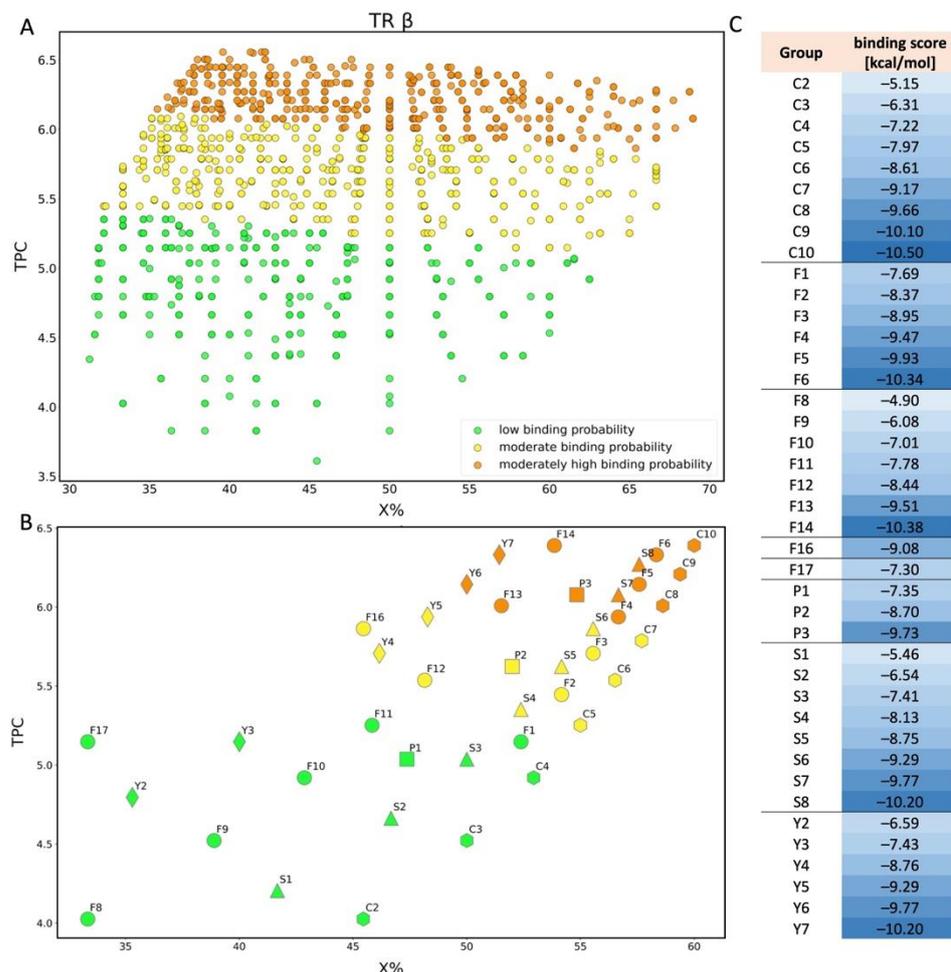


Figure 10. Scatter plot of binding scores obtained for TR β in X% and TPC space (A); scatter plot for TR β showing the division of PFAS into 5 different structural groups in space of TPC and X%. C-carboxylic acids (COOH-(CF₂)_n-CF₃), S-sulfonic acids (SO₃H-(CF₂)_n-CF₃), P-phosphonic acids (PO₃H₂-(CF₂)_n-CF₃, F-Fluorotelomer alcohol (F1–F7 (OH-(CH₂)-(CF₂)_n-CF₃), F8–F14 (OH-(CH₂)₂-(CF₂)_n-CF₃), F16 (OH-(CH₂)₃-(CF₂)_n-CF₃), F17 (OH-(CH₂)₄-(CF₂)_n-CF₃)), Y-dicarboxylic acids (COOH-(CF₂)_n-COOH). Colors: green-low binding probability, yellow-moderate binding probability, orange-moderately high binding probability (B); table showing binding score values for each compound shown in chart B (C).

3. Discussion—The Biological Significance of the In Silico Predictions

In the presented work we raised the important topic regarding the effects of the presence of PFAS in the environment and their impact on humans. Due to the constantly emerging PFAS in the environment, it is crucial to evaluate various perfluorinated compounds in terms of disrupting the human endocrine system. This study documents that the chemical structure of PFAS (particularly the carbon chain length, the number of fluorine atoms in a structure, and the number of branches) may have an influence on binding to the selected human nuclear hormone receptors. Therefore, the use of molecular docking in conjunction with the QSAR model is needed.

3.1. Peroxisome Proliferator-Activated Receptors

It is generally recognized that all PPAR isoforms bind to DNA as heterodimers with retinoid X-receptor (RXR), therefore PPAR/RXR heterodimers can be activated by RXR ligands as well [68]. However, our simulations were based solely on PFAS interactions with particular PPAR isoforms, more specifically with the PPAR ligand binding domain (LBD). In general, our analysis indicates that activation of PPAR α , β , or γ increases with an increasing chain length of carbons saturated with fluorine. This suggests that the hydrophobic interactions between the carbon chains of PFAS and amino acid residues in the binding pocket contribute significantly to the stabilization of the PFAS/protein complex [59]. The analysis shows that the considered PFAS group bind to the PPAR α , β , and γ only with low or moderate probability. The fact that all the analyzed PFAS showed only low binding strength to PPAR α (in contrast to PPAR β and PPAR γ for which some PFAS showed moderate binding probability) could be explained by slight differences in amino acid sequence and spatial conformation in the LBD. The molecular analysis of all three isoforms leads to rather surprising conclusions. As was proved by Xu et al [69], PPAR α isoform is composed of a helical sandwich and a four-stranded β -sheet, similar to PPAR β and PPAR γ . However, the PPAR α and PPAR γ LBD are significantly larger and more resemble each other in terms of size and shape in comparison to the PPAR β pocket which has a narrowing of the pocket adjacent to the activation function-2 (AF-2) surface. One can assume that such a situation should favor easier binding of larger variety of PFAS molecules, especially those of bigger sizes, to PPAR α and PPAR γ in comparison to PPAR β . It remains to be investigated why we observed comparatively lower binding of PFAS to PPAR α . The one difference in the case of the PPAR α pocket is that it is considered more lipophilic and less solvent exposed than the LBD of PPAR β and PPAR γ [69]. Possibly, the more hydrophobic nature of the PPAR α pocket may predispose it to bind to the more lipophilic PFAS.

Our *in silico* predictions may be in contrast to the biological results obtained by other researchers. Generally, PFAS were shown to activate mouse and human PPAR α in the *in vitro* assays using cells transfected with PPAR α reporter constructs, with higher sensitivity of the mouse assays [22–24]. In these studies, the PFAS with a carboxylic acid group showed a higher potential to activate PPAR α than the PFAS with a sulfonic acid group [22]. PFAS potential towards activating PPAR β and PPAR γ also differed depending on the carbon chain length and the terminal functional group. Interestingly, in the study by Li et al. [58] the three perfluorinated sulfonic acids used exhibited a stronger binding potency to PPAR β than their carboxylic counterparts, a phenomenon which is opposite to the results generally observed for PPAR α . These observations indicate rather complex and yet not well understood interactions of PFAS with PPARs. It is known that the *in vitro* assays have their limitations and do not necessarily reflect complex nuclear receptor interactions *in vivo*, e.g., activation of other receptors by ligands, the influence of co-activators, or endogenous ligands. In our opinion, while assessing the activating effects of PFAS on PPARs (but also on other NHRs) the possibility of their binding to alternate sites not only to the canonical LBD should be considered [70].

3.2. Thyroid Hormone Receptors

In the case of TR α and TR β , we observed that both the shape and size of the PFAS molecule or the percentage of fluorine atoms have a considerable influence on binding strength to these receptors. The PFAS compounds with longer chains showed a moderately high probability of binding. Our results are to some extent corroborated by the study of Mortensen et al., who predicted that in the gull TR α model several PFAS, including two perfluoroalkyl carboxylic acids (PFCA), could bind to the LBD [71]. In that study, long-chained PFCA were the strongest binders, possibly due to their enhanced hydrophobicity and the highly hydrophobic LBD of TR α . Based on the data some PFASs that are not aromatic and have a different structure from the thyroid hormones, resembling more fatty acids and acyl-CoA esters, can be predicted to be relatively good TR ligands. PFOS

was predicted in our study as a good binder to TR α and TR β , which was supported by molecular docking and biological testing by Ren et al. [72] and Xiu et al. [73]. According to Xiu et al., PFOS fits into the LBD of TR α and TR β with the negative charged end group bound near the entrance of the LBD, and the hydrophobic part towards the interior of the LBD [73].

In our study perfluoroacids with C8–C10 were classified as having a moderately high binding potential to both TRs, which was fully supported by the results obtained by Ren et al. [72] who reported that acids with C8–C18 effectively bound to TR α -LBD. Moreover, a very interesting correlation between our predictions and the biological relative potency of binding of perfluoroalkyl sulfonates to TR α -LBD as reported in the study by Ren et al. [72] can be observed. In the case of TR α , we observed an increased binding probability for sulfonic acids with ≥ 5 carbon atoms in the structure (S4–S8). A very similar increasing trend was observed in the TR α -LBD binding assay by Ren et al. [72] for perfluorobutane sulfonate (PFBS, RP = 0.0003 relative to T3), perfluorohexane sulfonate (PFHxS, RP = 0.0015 relative to T3), perfluorooctane sulfonate (PFOS, RP = 0.01875 relative to T3).

4. Materials and Methods

4.1. Dataset

Dataset for the development QSAR models.

The commonly used 43 PFAS encompassing perfluorinated carboxylic acids (PFCAs), fluorotelomer alcohols (FTOHs), perfluorinated sulfonic acids (PFASs), and other related PFAS, were selected [19]. Names, acronyms, CAS numbers, and SMILES of PFAS examined in this study are listed in Supplementary Materials S1.

Dataset for virtual screening of PFAS.

The external dataset including more than 5000 per- and polyfluoroalkyl compounds from NORMAN Database System was downloaded [57]. The chemicals without SMILES were removed from the PFAS dataset. The resultant dataset contains a total of 4464 PFAS. For all details regarding this dataset please refer to Supplementary Materials S1.

4.2. Docking Score Calculations

Docking scores of 43 PFAS (using SMILES code) to five human nuclear hormone receptors (PPAR α , PPAR β , PPAR γ , TR α , TR β) were calculated using the Endocrine Disruptome-open-source prediction tool [41] based on the technique of inverse docking. The 3D crystal structures integrated with Endocrine Disruptome are 3KDU (PPAR α), 3GZ9 (PPAR β), 3ET3 (PPAR γ), 3ILZ (TR α), 3IMY (TR β). All modeled compounds have a molecular weight lower than 600 g/mol, which is in line with the accuracy of the predictions of this tool. Docking simulations are provided by Docking Interface for Target Systems (DOTS) [41] platform, while the docking calculations were run with AutoDock Vina.

Obtained results were color-coded into four probability binding classes according to the thresholds set in the tool [41], where “green” corresponds to a low probability of binding, “yellow” corresponds to a moderate probability of binding, “orange” indicates a moderately high binding probability and “red” indicates a high probability of binding (Table 2). For docking results please refer to Supplementary Materials S1.

Table 2. Threshold values of binding free energies for PPAR α , PPAR β , PPAR γ , TR α , TR β .

Nuclear Hormone Receptor	High probability	Threshold 1	Moderately high probability	Threshold 2	Moderate probability	Threshold 3	Low probability
PPAR α		−10		−9.4		−8.9	
PPAR β		−10.5		−10.1		−9.6	
PPAR γ		−10.3		−9.6		−8.9	
TR α		−10.2		−9.2		−7.2	
TR β		−10.5		−9.4		−7.8	

4.3. Molecular Descriptors

The molecular descriptors are chemical structures translated into numbers, which represent information about a chemical compound based on its chemical constitution elements. Based on the descriptors, it is possible to determine the correlation of chemical structure with biological activity [74]. For all selected 43 PFAS a total number of 186 1D and 2D descriptors were calculated in alvaDesc software (version 2.0.10, Alvascience, Lecco (Italy)) using SMILES notation [75]. These descriptors belong to constitutional indices, topological indices, walk and path counts, and molecular properties. The number of descriptors has been firstly reduced by constant values and by descriptors with no data for at least one compound. For the modeling, we applied the multiple linear regression method (MLR) combined with a genetic algorithm (GA). The GA was used to select the optimal combination of the previously calculated structural descriptors, to be utilized in the final model (Supplementary Materials S2).

4.4. QSAR Model Calibration and Validation

In this work five models to predict PFAS binding probability to PPAR α , PPAR β , PPAR γ , TR α , and TR β were developed. After collecting docking scores from the Endocrine Disruptome tool [41] the dataset was split into training (T) and validation (V) sets by following the “1:Z” algorithm [76]. In this case, every Zth compound in a group of chemicals, sorted by calculated values of docking score in ascending order, was assigned to the validation set, and the remaining compounds created the training set. Here, for PPAR α , PPAR β , and TR β , Z = 3, while for TR α and PPAR γ Z = 4, which means 14 and 10 compounds were assigned to the validation set, respectively. The training set is used to build the model and the validation set is necessary to assess the predictability of the QSAR model [77]. Before training the models for estimating binding energy to NHRs, the independent variables (descriptors) and dependent variables (docking scores) were preprocessed by standard scaling [78]. For scientific validity of the QSAR models [49] all of them were developed in accordance with the standards and recommendations of the Organization for Economic Co-operation and Development (OECD) [46–48]. The analyses were carried out using packages available in Python 3.8.8 [79]. For defining the relationship between the structure of PFAS and binding probability (endpoint) to NHRs the multiple linear regression (MLR) method [80,81] was applied. Structural descriptors, which have an influence on the probability of binding to the receptor, and which were readily interpretable, were selected. Only a combination of molecular descriptors, which correlation does not exceed 0.6, were considered. One of the most important stages in the calibration of the QSAR model is to verify the goodness-of-fit by calculating the appropriate parameters. In this work, the squared regression coefficient (R^2), which measures the variation of observed data with the fitted one (preferably if R^2 is equal to 1) and the root mean square error of calibration ($RMSE_C$), based on the prediction for the training set, were estimated. To demonstrate the robustness and predictive ability the validation of obtained models was divided into two main steps: (i) internal validation (by using the training set) and (ii) external validation (by using the validation set). In the case of internal validation, the leave-one-out cross-validation method was used. This involves modifying the original training set of n compounds by removing one compound and predicting its activity using the model developed with n-1 compounds. Several parameters were estimated for verifying the robustness of the developed models: cross-validation coefficient (Q^2_{LOO}), the root mean square error of cross-validation ($RMSE_{CV}$), mean absolute error (MAE_{CV}), concordance correlation coefficient (CCC_{CV}) [42]. The external validation of the developed models was performed using the validation set defined in the first step of the modeling procedure. To indicate if the developed models predict correctly, the external coefficients Q^2 (F1, F2, and F3), the root mean square error ($RMSE_{EXT}$), and the mean absolute error (MAE_{EXT}) were calculated. In addition, a parameter CCC_{EXT} (concordance correlation coefficient), which measures precision and accuracy, detecting the distance of the observations from the fitting line was obtained [82]. All calculated parameters can be found in Supplementary Materials

S1. A detailed description of the applied metrics can be found in the literature [83]. To ensure that developed models are not correlated by chance, the Y-randomization has been performed and the R^2Y_{SCR} was evaluated. The applicability domain (AD), defined as the scope and limitations of a model, was verified using the Williams plot [48,81] and Insubria graph [55] (according to h^*). Williams plot is a graphical representation of standardized residuals (y-axis) versus leverages (hat values; x-axis) for each compound of the dataset [84]. The QMRF documents for all five developed models are implemented in Supplementary Materials S2.

4.5. Screening of the Binding Potential to NHRs for a Large Set of PFAS

After the development and validation of models, they have been applied for predicting binding energies to peroxisome proliferator-activated (α , β , and γ) and thyroid hormone (α and β) receptors for the 4464 PFAS implemented in the NORMAN Database System, for which data are not available so far. All compounds with their Names, CAS numbers, and SMILES are attached in Supplementary Materials S1. The predictions were verified if they are in the AD using the Williams plot and Insubria Graph (involving leverage value h^*). Results obtained for all compounds were color coded into four classes according to the threshold set in the tool: [41] “green”—low probability of binding; “yellow”—moderately probability of binding; “orange”—moderately high binding probability; and “red”—a high probability of binding (Table 2). Screening of a new comprehensive dataset was conducted to determine the dependence of receptor binding strength on a chemical structure.

5. Conclusions

As a result of intensive industrial development, new diverse PFAS are constantly emerging in the environment. Therefore, it is important to implement *in silico* methods to assess their impact on the human body. A combination of molecular docking and QSAR models makes it possible to determine which structural features of a chemical compound influence endocrine system disruption. In this work, we have presented an analysis of the applicability of the developed models, which shows that they can be successfully used to predict the probability of PFAS binding to nuclear human receptors. Furthermore, it should be borne in mind that compounds with very high structural variability appear in the environment, so they should be approached with caution.

When analyzing the classification of PFAS into functional groups, the trend was observed indicating an increase in binding strength to the selected NHRs as the descriptor values increase. The analysis of available literature data on biological significance of the observed predictions generally correlated well with the measured binding potential of PFAS to the NHRs in biological assays. The differences observed may result from some limitations of the *in vitro* assays, mainly the possibility of activating NHRs through the binding of PFAS to alternative binding sites in NHRs. In our opinion compounds from the NORMAN Database include various specific elements and functional groups, therefore the developed models based on selected molecular descriptors are rather suitable for linear perfluorinated compounds. To gain additional knowledge on PFAS binding to human nuclear receptors, further scientific studies are needed. However, our results suggest that PPAR α,β,γ and TR α,β mediated pathways should be taken into consideration when studying the toxicity mechanisms of PFAS.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/molecules28020479/s1>, QMRF prepared for developed models.

Author Contributions: T.P. and A.S. formulated the research problem and main hypotheses. A.S. designed the study. D.K., A.S. and N.B. conducted the research. A.S., D.K., N.B., M.S., H.B., P.B. and T.P. contributed to the analysis and discussion. A.S., N.B., D.K. and M.S. wrote the manuscript. All authors edited the manuscript and accepted its final version. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N°101036449.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare no conflict of interest.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. Kissa, E. *Fluorinated Surfactants and Repellents*; Surfactant Science Series; Marcel Dekker: New York, NY, USA, 2001; Volume 97.
2. Glüge, J.; Scheringer, M.; Cousins, I.T.; DeWitt, J.C.; Goldenman, G.; Herzke, D.; Lohmann, R.; Ng, C.A.; Trier, X.; Wang, Z. An Overview of the Uses of Per- and Polyfluoroalkyl Substances (PFAS). *Environ. Sci. Process Impacts* **2020**, *22*, 2345–2373. [[CrossRef](#)]
3. OECD. *Per-and Polyfluoroalkyl Substances and Alternatives in Coatings, Paints and Varnishes (CPVs), Report on the Commercial Availability and Current Uses*; OECD Series on Risk Management, No. 70, Environment, Health and Safety, Environment Directorate; OECD: Paris, France, 2022.
4. Prevedouros, K.; Cousins, I.T.; Buck, R.C.; Korzeniowski, S.H. Sources, Fate and Transport of Perfluorocarboxylates. *Environ. Sci. Technol.* **2005**, *40*, 32–44. [[CrossRef](#)] [[PubMed](#)]
5. Pelch, K.E.; Reade, A.; Kwiatkowski, C.F.; Merced-Nieves, F.M.; Cavalier, H.; Schultz, K.; Wolffe, T.; Varshavsky, J. The PFAS-Tox Database: A Systematic Evidence Map of Health Studies on 29 per- and Polyfluoroalkyl Substances. *Environ. Int.* **2022**, *167*, 107408. [[CrossRef](#)] [[PubMed](#)]
6. Boyd, R.I.; Ahmad, S.; Singh, R.; Fazal, Z.; Prins, G.S.; Erdogan, Z.M.; Irudayaraj, J.; Spinella, M.J. Toward a Mechanistic Understanding of Poly- and Perfluoroalkylated Substances and Cancer. *Cancers* **2022**, *14*, 2919. [[CrossRef](#)] [[PubMed](#)]
7. Boulanger, B.; Vargo, J.D.; Schnoor, J.L.; Hornbuckle, K.C. Evaluation of Perfluorooctane Surfactants in a Wastewater Treatment System and in a Commercial Surface Protection Product. *Environ. Sci. Technol.* **2005**, *39*, 5524–5530. [[CrossRef](#)]
8. Harada, K.; Nakanishi, S.; Sasaki, K.; Furuyama, K.; Nakayama, S.; Saito, N.; Yamakawa, K.; Koizumi, A. Particle Size Distribution and Respiratory Deposition Estimates of Airborne Perfluorooctanoate and Perfluorooctanesulfonate in Kyoto Area, Japan. *Bull. Environ. Contam. Toxicol.* **2006**, *76*, 306–310. [[CrossRef](#)]
9. Emmett, E.A.; Zhang, H.; Shofer, F.S.; Freeman, D.; Rodway, N.V.; Desai, C.; Shaw, L.M. Community Exposure to Perfluorooctanoate: Relationships Between Serum Levels and Certain Health Parameters. *J. Occup. Environ. Med.* **2006**, *48*, 771–779. [[CrossRef](#)]
10. Butenhoff, J.L.; Gaylor, D.W.; Moore, J.A.; Olsen, G.W.; Rodricks, J.; Mandel, J.H.; Zobel, L.R. Characterization of Risk for General Population Exposure to Perfluorooctanoate. *Regul. Toxicol. Pharm.* **2004**, *39*, 363–380. [[CrossRef](#)]
11. Calafat, A.M.; Wong, L.-Y.; Kuklenyik, Z.; Reidy, J.A.; Needham, L.L. Polyfluoroalkyl Chemicals in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 2003–2004 and Comparisons with NHANES 1999–2000. *Environ. Health Persp.* **2007**, *115*, 1596–1602. [[CrossRef](#)]
12. Kudo, N.; Suzuki-Nakajima, E.; Mitsumoto, A.; Kawashima, Y. Responses of the Liver to Perfluorinated Fatty Acids with Different Carbon Chain Length in Male and Female Mice: In Relation to Induction of Hepatomegaly, Peroxisomal β -Oxidation and Microsomal 1-Acylglycerophosphocholine Acyltransferase. *Biol. Pharm. Bull.* **2006**, *29*, 1952–1957. [[CrossRef](#)]
13. Kärrman, A.; Ericson, I.; van Bavel, B.; Darnerud, P.O.; Aune, M.; Glynn, A.; Lignell, S.; Lindström, G. Exposure of Perfluorinated Chemicals through Lactation: Levels of Matched Human Milk and Serum and a Temporal Trend, 1996–2004, in Sweden. *Environ. Health Persp.* **2007**, *115*, 226–230. [[CrossRef](#)] [[PubMed](#)]
14. Frawley, R.P.; Smith, M.; Cesta, M.F.; Hayes-Bouknight, S.; Blystone, C.; Kissling, G.E.; Harris, S.; Germolec, D. Immunotoxic and Hepatotoxic Effects of Perfluoro-n-Decanoic Acid (PFDA) on Female Harlan Sprague-Dawley Rats and B6C3F1/N Mice When Administered by Oral Gavage for 28 Days. *J. Immunotoxicol.* **2018**, *15*, 41–52. [[CrossRef](#)] [[PubMed](#)]
15. Innes, K.E.; Wimsatt, J.H.; Frisbee, S.; Ducatman, A.M. Inverse Association of Colorectal Cancer Prevalence to Serum Levels of Per-fluorooctane Sulfonate (PFOS) and Perfluorooctanoate (PFOA) in a Large Appalachian Population. *BMC Cancer* **2014**, *14*, 45. [[CrossRef](#)] [[PubMed](#)]
16. Gaballah, S.; Swank, A.; Sobus, J.R.; Howey, X.M.; Schmid, J.; Catron, T.; McCord, J.; Hines, E.; Strynar, M.; Tal, T. Evaluation of Developmental Toxicity, Developmental Neurotoxicity, and Tissue Dose in Zebrafish Exposed to GenX and Other PFAS. *Environ. Health Persp.* **2020**, *128*, 47005. [[CrossRef](#)]
17. SC-4/17: Perfluorooctane Sulfonic Acid, Its Salts and Perfluorooctane Sulfonyl Fluoride. In Proceedings of the Fourth Meeting of the Conference of the Parties to the Stockholm Convention, Geneva, Switzerland, 4–8 May 2009. Available online: <http://Chm.Pops.Int/TheConvention/ThePOPs/TheNewPOPs/Tabid/2511/Default.Asp?L> (accessed on 2 November 2022).

18. SC-9/12: Perfluorooctanoic Acid (PFOA), Its Salts and PFOA-Related Compounds. In Proceedings of the Ninth Meeting of the Conference of the Parties to the Stockholm Convention, Geneva, Switzerland, 29 April–10 May 2019. Available online: <http://Chm.Pops.Int/TheConvention/ThePOPs/TheNewPOPs/Tabid/2511/Default.aspx> (accessed on 2 November 2022).
19. Yu, S.; Ren, J.; Lv, Z.; Li, R.; Zhong, Y.; Yao, W.; Yuan, J. Prediction of the Endocrine-Disrupting Ability of 49 per- and Polyfluoroalkyl Substances: In Silico and Epidemiological Evidence. *Chemosphere* **2022**, *290*, 133366. [[CrossRef](#)] [[PubMed](#)]
20. Cheng, W.; Ng, C.A. Using Machine Learning to Classify Bioactivity for 3486 Per- and Polyfluoroalkyl Substances (PFASs) from the OECD List. *Environ. Sci. Technol.* **2019**, *53*, 13970–13980. [[CrossRef](#)]
21. Walker, C.; Garza, S.; Papadopoulos, V.; Culty, M. Impact of Endocrine-Disrupting Chemicals on Steroidogenesis and Consequences on Testicular Function. *Mol. Cell Endocrinol.* **2021**, *527*, 111215. [[CrossRef](#)]
22. Behr, A.-C.; Plinsch, C.; Braeuning, A.; Buhrke, T. Activation of Human Nuclear Receptors by Perfluoroalkylated Substances (PFAS). *Toxicol. In Vitro* **2020**, *62*, 104700. [[CrossRef](#)]
23. Takacs, M.L.; Abbott, B.D. Activation of Mouse and Human Peroxisome Proliferator-Activated Receptors (α , β/δ , γ) by Perfluorooctanoic Acid and Perfluorooctane Sulfonate. *Toxicol. Sci.* **2007**, *95*, 108–117. [[CrossRef](#)]
24. Wolf, C.J.; Takacs, M.L.; Schmid, J.E.; Lau, C.; Abbott, B.D. Activation of Mouse and Human Peroxisome Proliferator-Activated Receptor Alpha by Perfluoroalkyl Acids of Different Functional Groups and Chain Lengths. *Toxicol. Sci.* **2008**, *106*, 162–171. [[CrossRef](#)]
25. Christofides, A.; Konstantinidou, E.; Jani, C.; Boussiotis, V.A. The Role of Peroxisome Proliferator-Activated Receptors (PPAR) in Immune Responses. *Metabolism* **2020**, *114*, 154338. [[CrossRef](#)] [[PubMed](#)]
26. Chawla, A.; Repa, J.J.; Evans, R.M.; Mangelsdorf, D.J. Nuclear Receptors and Lipid Physiology: Opening the X-Files. *Science* **2001**, *294*, 1866–1870. [[CrossRef](#)] [[PubMed](#)]
27. Berger, J.; Moller, D.E. The Mechanisms of Action of PPARs. *Annu. Rev. Med.* **2002**, *53*, 409–435. [[CrossRef](#)] [[PubMed](#)]
28. Shipley, J.M.; Hurst, C.H.; Tanaka, S.S.; DeRoos, F.L.; Butenhoff, J.L.; Seacat, A.M.; Waxman, D.J. Trans-Activation of PPAR α and Induction of PPAR α Target Genes by Perfluorooctane-Based Chemicals. *Toxicol. Sci.* **2004**, *80*, 151–160. [[CrossRef](#)] [[PubMed](#)]
29. Heuvel, J.P.V.; Thompson, J.T.; Frame, S.R.; Gillies, P.J. Differential Activation of Nuclear Receptors by Perfluorinated Fatty Acid Analogs and Natural Fatty Acids: A Comparison of Human, Mouse, and Rat Peroxisome Proliferator-Activated Receptor- α , - β , and - γ , Liver X Receptor- β , and Retinoid X Receptor- α . *Toxicol. Sci.* **2006**, *92*, 476–489. [[CrossRef](#)]
30. Tao, L.J.; Seo, D.E.; Jackson, B.; Ivanova, N.B.; Santori, F.R. Nuclear Hormone Receptors and Their Ligands: Metabolites in Control of Transcription. *Cells* **2020**, *9*, 2606. [[CrossRef](#)]
31. Eliceiri, B.P.; Brown, D.D. Quantitation of Endogenous Thyroid Hormone Receptors Alpha and Beta during Embryogenesis and Metamorphosis in *Xenopus Laevis*. *J. Biol. Chem.* **1994**, *269*, 24459–24465. [[CrossRef](#)]
32. Collet, B.; Simon, E.; van der Linden, S.; El Abdellaoui, N.; Naderman, M.; Man, H.; Middelhof, I.; van der Burg, B.; Besselink, H.; Brouwer, A. Evaluation of a Panel of in Vitro Methods for Assessing Thyroid Receptor β and Transthyretin Transporter Disrupting Activities. *Reprod. Toxicol.* **2020**, *96*, 432–444. [[CrossRef](#)]
33. Jabbar, A.; Pingitore, A.; Pearce, S.H.S.; Zaman, A.; Iervasi, G.; Razvi, S. Thyroid Hormones and Cardiovascular Disease. *Nat. Rev. Cardiol.* **2016**, *14*, 39–55. [[CrossRef](#)]
34. Szilagyi, J.T.; Avula, V.; Fry, R.C. Perfluoroalkyl Substances (PFAS) and Their Effects on the Placenta, Pregnancy, and Child Development: A Potential Mechanistic Role for Placental Peroxisome Proliferator-Activated Receptors (PPARs). *Curr. Environ. Health Rep.* **2020**, *7*, 222–230. [[CrossRef](#)]
35. Jones, P.D.; Hu, W.; Coen, W.D.; Newsted, J.L.; Giesy, J.P. Binding of Perfluorinated Fatty Acids to Serum Proteins. *Environ. Toxicol. Chem.* **2003**, *22*, 2639–2649. [[CrossRef](#)] [[PubMed](#)]
36. Han, X.; Nabb, D.L.; Russell, M.H.; Kennedy, G.L.; Rickard, R.W. Renal Elimination of Perfluorocarboxylates (PFCAs). *Chem. Res. Toxicol.* **2012**, *25*, 35–46. [[CrossRef](#)] [[PubMed](#)]
37. Shrestha, S.; Bloom, M.S.; Yucel, R.; Seegal, R.F.; Wu, Q.; Kannan, K.; Rej, R.; Fitzgerald, E.F. Perfluoroalkyl Substances and Thyroid Function in Older Adults. *Environ. Int.* **2015**, *75*, 206–214. [[CrossRef](#)] [[PubMed](#)]
38. Wyrzykowska, E.; Rybińska-Fryca, A.; Sosnowska, A.; Puzyn, T. Virtual Screening in the Design of Ionic Liquids as Environmentally Safe Bactericides. *Green Chem.* **2019**, *21*, 1965–1973. [[CrossRef](#)]
39. Barycki, M.; Sosnowska, A.; Puzyn, T. AquaBoxIL—A Computational Tool for Determining the Environmental Distribution Profile of Ionic Liquids. *Green Chem.* **2018**, *20*, 3359–3370. [[CrossRef](#)]
40. Ahuja, S.; Deep, P.; Shravani, N.; Sambhyal, S.; Mishra, D.; Pandey, C.; Manchanda, P.; Krishna, D.; Dee, A.; et al. Molecular Docking; Future of Medicinal Research. *Ecol. Environ. Conserv.* **2022**, *28*, 18. [[CrossRef](#)]
41. Kolsek, K.; Mavri, J.; Dolenc, M.S.; Gobec, S.; Turk, S. Endocrine Disruptome—An Open Source Prediction Tool for Assessing Endocrine Disruption Potential through Nuclear Receptor Binding. *J. Chem. Inf. Model* **2014**, *54*, 1254–1267. [[CrossRef](#)]
42. Sangion, A.; Gramatica, P. PBT Assessment and Prioritization of Contaminants of Emerging Concern: Pharmaceuticals. *Environ. Res.* **2016**, *147*, 297–306. [[CrossRef](#)]
43. Puzyn, T.; Leszczynska, D.; Leszczynski, J. Toward the Development of “Nano-QSARs”: Advances and Challenges. *Small Weinheim. Bergstr. Ger.* **2009**, *5*, 2494–2509. [[CrossRef](#)]
44. Gramatica, P. Recent Advances in QSAR Studies. *Chall. Adv. Comput. Chem. Phys.* **2009**, *8*, 327–366. [[CrossRef](#)]

45. Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T.P.; Michalkova, A.; Hwang, H.-M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using Nano-QSAR to Predict the Cytotoxicity of Metal Oxide Nanoparticles. *Nat. Nanotechnol.* **2011**, *6*, 175–178. [[CrossRef](#)] [[PubMed](#)]
46. OECD. OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure Activity Relationship Models. In *37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology*; Organisation for Economic Co-Operation and Development: Paris, France, 2004.
47. OECD. *Guidance Document on the Validation of (Quantitative) Structure Activity Relationships [QSAR] Models*; Organisation for Economic Co-Operation and Development: Paris, France, 2007.
48. Tropsha, A.; Gramatica, P.; Gombar, V. The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *Qsar. Comb. Sci.* **2003**, *22*, 69–77. [[CrossRef](#)]
49. Sosnowska, A.; Bulawska, N.; Kowalska, D.; Puzyn, T. Towards Higher Scientific Validity and Regulatory Acceptance of Predictive Models for PFAS. 2022; *manuscript submitted for publication*.
50. Deshpande, S.; Solomon, V.R.; Katti, S.B.; Prabhakar, Y.S. Topological Descriptors in Modelling Antimalarial Activity: N(1)-(7-Chloro-4-Quinolyl)-1,4-Bis(3-Aminopropyl)Piperazine as Prototype. *J. Enzym. Inhib. Med. Chem.* **2009**, *24*, 94–104. [[CrossRef](#)]
51. da Veras, L.S.; Arakawa, M.; Funatsu, K.; Takahata, Y. 2D and 3D QSAR Studies of the Receptor Binding Affinity of Progestins. *J. Brazil Chem. Soc.* **2010**, *21*, 872–881. [[CrossRef](#)]
52. Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.T.D.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; et al. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. *Altern Lab. Anim.* **2005**, *33*, 155–173. [[CrossRef](#)]
53. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structures (Vol. 5)*; Research Studies Press: Chichester, UK, 1983.
54. Li, F.; Zhang, C.; Qu, Y.; Chen, J.; Chen, L.; Liu, Y.; Zhou, Q. Quantitative Characterization of Short- and Long-Chain Perfluorinated Acids in Solid Matrices in Shanghai, China. *Sci. Total Environ.* **2009**, *408*, 617–623. [[CrossRef](#)]
55. Gramatica, P.; Cassani, S.; Roy, P.P.; Kovarich, S.; Yap, C.W.; Papa, E. QSAR Modeling Is Not “Push a Button and Find a Correlation”: A Case Study of Toxicity of (Benzo-)Triazoles on Algae. *Mol. Inform.* **2012**, *31*, 817–835. [[CrossRef](#)]
56. Put, R.; Perrin, C.; Questier, F.; Coomans, D.; Massart, D.L.; Heyden, Y.V. Classification and Regression Tree Analysis for Molecular Descriptor Selection and Retention Prediction in Chromatographic Quantitative Structure–Retention Relationship Studies. *J. Chromatogr. A* **2003**, *988*, 261–276. [[CrossRef](#)]
57. Norman Network: Norman Substance Database (SusDat). Available online: <https://www.Norman-network.Com/Nds/Susdat/> (accessed on 17 May 2022).
58. Li, C.-H.; Ren, X.-M.; Cao, L.-Y.; Qin, W.-P.; Guo, L.-H. Investigation of Binding and Activity of Perfluoroalkyl Substances to the Human Peroxisome Proliferator-Activated Receptor β/δ . *Environ. Sci. Process Impacts* **2019**, *21*, 1908–1914. [[CrossRef](#)]
59. Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol* **2013**, *47*, 11293–11301. [[CrossRef](#)]
60. Xie, W.; Zhong, W.; Appenzeller, B.M.R.; Zhang, J.; Junaid, M.; Xu, N. Nexus between Perfluoroalkyl Compounds (PFCs) and Human Thyroid Dysfunction: A Systematic Review Evidenced from Laboratory Investigations and Epidemiological Studies. *Crit. Rev. Environ. Sci. Tec.* **2020**, *51*, 2485–2530. [[CrossRef](#)]
61. Available online: <https://www.Rcsb.Org/Structure/3ILZ> (accessed on 2 November 2022).
62. Available online: <https://www.Rcsb.Org/Structure/3IMY> (accessed on 2 November 2022).
63. Yao, B.; Yang, C.; Pan, C.; Li, Y. Thyroid Hormone Resistance: Mechanisms and Therapeutic Development. *Mol. Cell Endocrinol.* **2022**, *553*, 111679. [[CrossRef](#)] [[PubMed](#)]
64. Coperchini, F.; Croce, L.; Pignatti, P.; Ricci, G.; Gangemi, D.; Magri, F.; Imbriani, M.; Rotondi, M.; Chiovato, L. The New Generation PFAS C6O4 Does Not Produce Adverse Effects on Thyroid Cells in Vitro. *J. Endocrinol. Investig.* **2021**, *44*, 1625–1635. [[CrossRef](#)] [[PubMed](#)]
65. Coperchini, F.; Croce, L.; Ricci, G.; Magri, F.; Rotondi, M.; Imbriani, M.; Chiovato, L. Thyroid Disrupting Effects of Old and New Generation PFAS. *Front. Endocrinol.* **2021**, *11*, 612320. [[CrossRef](#)] [[PubMed](#)]
66. Young, A.S.; Zoeller, T.; Hauser, R.; James-Todd, T.; Coull, B.A.; Behnisch, P.A.; Brouwer, A.; Zhu, H.; Kannan, K.; Allen, J.G. Assessing Indoor Dust Interference with Human Nuclear Hormone Receptors in Cell-Based Luciferase Reporter Assays. *Environ. Health Persp.* **2021**, *129*, 47010. [[CrossRef](#)] [[PubMed](#)]
67. Behnisch, P.A.; Besselink, H.; Weber, R.; Willand, W.; Huang, J.; Brouwer, A. Developing Potency Factors for Thyroid Hormone Disruption by PFASs Using TTR-TR β CALUX[®] Bioassay and Assessment of PFASs Mixtures in Technical Products. *Environ. Int.* **2021**, *157*, 106791. [[CrossRef](#)]
68. Metzger, D.; Imai, T.; Jiang, M.; Takukawa, R.; Desvergne, B.; Wahli, W.; Chambon, P. Functional Role of RXRs and PPARgamma in Mature Adipocytes. *Prostaglandins Leukot. Essent. Fat. Acids* **2005**, *73*, 51–58. [[CrossRef](#)]
69. Xu, H.E.; Lambert, M.H.; Montana, V.G.; Plunket, K.D.; Moore, L.B.; Collins, J.L.; Oplinger, J.A.; Klierer, S.A.; Gampe, R.T.; McKee, D.D.; et al. Structural Determinants of Ligand Binding Selectivity between the Peroxisome Proliferator-Activated Receptors. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13919–13924. [[CrossRef](#)]
70. Hughes, T.S.; Giri, P.K.; de Vera, I.M.S.; Marciano, D.P.; Kuruvilla, D.S.; Shin, Y.; Blayo, A.-L.; Kamenecka, T.M.; Burris, T.P.; Griffin, P.R.; et al. An Alternate Binding Site for PPAR γ Ligands. *Nat. Commun.* **2014**, *5*, 3571. [[CrossRef](#)]

71. Mortensen, Å.-K.; Mæhre, S.; Kristiansen, K.; Heimstad, E.S.; Gabrielsen, G.W.; Jenssen, B.M.; Sylte, I. Homology Modeling to Screen for Potential Binding of Contaminants to Thyroid Hormone Receptor and Transthyretin in Glaucous Gull (*Larus Hyperboreus*) and Herring Gull (*Larus Argentatus*). *Comput. Toxicol.* **2020**, *13*, 100120. [CrossRef]
72. Ren, X.-M.; Zhang, Y.-F.; Guo, L.-H.; Qin, Z.-F.; Lv, Q.-Y.; Zhang, L.-Y. Structure–Activity Relations in Binding of Perfluoroalkyl Compounds to Human Thyroid Hormone T3 Receptor. *Arch. Toxicol.* **2015**, *89*, 233–242. [CrossRef] [PubMed]
73. Xin, Y.; Ren, X.-M.; Ruan, T.; Li, C.-H.; Guo, L.-H.; Jiang, G. Chlorinated Polyfluoroalkylether Sulfonates Exhibit Similar Binding Potency and Activity to Thyroid Hormone Transport Proteins and Nuclear Receptors as Perfluorooctanesulfonate. *Environ. Sci. Technol.* **2018**, *52*, 9412–9418. [CrossRef] [PubMed]
74. Consonni, V.; Todeschini, R. *Molecular Descriptors Recent Advances in QSAR Studies*; Springer: New York, NY, USA, 2010; pp. 29–102.
75. Mauri, A. *Ecotoxicological QSARs; Methods in Pharmacology and Toxicology*; Humana: New York, NY, USA, 2020; pp. 801–820. [CrossRef]
76. Puzyn, T.; Mostrag-Szlichtyng, A.; Gajewicz, A.; Skrzyński, M.; Worth, A.P. Investigating the Influence of Data Splitting on the Pre-dictive Ability of QSAR/QSPR Models. *Struct. Chem.* **2011**, *22*, 795–804. [CrossRef]
77. De, P.; Kar, S.; Ambure, P.; Roy, K. Prediction Reliability of QSAR Models: An Overview of Various Validation Tools. *Arch. Toxicol.* **2022**, *96*, 1279–1295. [CrossRef]
78. Available online: <https://Scikit-Learn.Org/Stable/Modules/Generated/Sklearn.Preprocessing.StandardScaler.Html> (accessed on 2 November 2022).
79. Python Software Foundation. Python Language Reference, Version 3.8.8. Available online: <http://Www.Python.Org> (accessed on 2 November 2022).
80. Pires, A.M.; Rodrigues, I.M. Multiple Linear Regression with Some Correlated Errors: Classical and Robust Methods. *Stat. Med.* **2007**, *26*, 2901–2918. [CrossRef]
81. Gramatica, P. Principles of QSAR Models Validation: Internal and External. *Qsar. Comb Sci.* **2007**, *26*, 694–701. [CrossRef]
82. Chirico, N.; Gramatica, P. 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* **2011**, *51*, 2320–2335. [CrossRef]
83. Alexander, D.L.J.; Tropsha, A.; Winkler, D.A. Beware of R 2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model* **2015**, *55*, 1316–1322. [CrossRef]
84. Helguera, A.; Cordeiro, M.N.; González, M.; Pérez, M.C.; Ruiz, R.; Castillo, Y. QSAR Modeling for Predicting Carcinogenic Potency of Nitroso-Compounds Using 0D-2D Molecular Descriptors. In Proceedings of the 11th International Electronic Conference on Synthetic Organic Chemistry, Basel, Switzerland, 1–30 November 2007; p. 1361. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.