UNIVERSITY OF INSUBRIA

PhD Program in Chemical and Environmental Sciences

Cycle XXXVIII

Non-Animal Approaches to Address Thyroid Hormone System Disruption

By

Marco Evangelista

Supervisor: Prof. Ester Papa, PhD
Co-supervisor: Dr. Nicola Chirico, PhD

Varese-Como 2025

Table of contents

Abstract

This PhD project addressed the critical need for a more efficient and ethical approach to chemical safety assessment, aligning with the EU's Chemicals Strategy for Sustainability (CSS) and the requirement to reduce reliance on (vertebrate) animal-based testing and to identify the most harmful chemical substances. The research focused on advancing new approach methodologies (NAMs), specifically quantitative structure-activity relationship (QSAR) models, for evaluating chemicals able to disrupt the thyroid hormone (TH) system and consequently lead to a wide range of severe adverse effects, such as neurotoxicity.

The central aim was the development of novel QSAR models to predict the ability of chemical substances to bind to human transthyretin (hTTR), a key molecular initiating event (MIE) in adverse outcome pathways (AOPs) for TH system disruption. This involved creating the largest, most comprehensive dataset of hTTR binding values collected and consolidated from the literature at that time, leading to the development of three new robust, *in vitro* assay-specific regression QSAR models. Furthermore, dedicated models were developed for per- and polyfluoroalkyl substances (PFAS), including a classification QSAR to first identify potential hTTR disruptors and a regression QSAR to quantify their T4-hTTR competing potencies. Overall, a strong focus was placed on the mechanistic interpretation of the selected molecular descriptors.

A comprehensive review of existing QSAR models for TH system disruption was performed to define the state-of-the-art. This foundational objective was dedicated to QSAR models predicting relevant MIEs to meet the needs of integrating NAMs within the AOP framework for TH system disruption assessment. The analysis systematically identified knowledge gaps, providing a clear foundation for future research priorities.

Crucially, the new QSARs were implemented into the open-access QSAR-ME Profiler software to facilitate their use, promoting transparency, and providing robust quantification of predictions uncertainty beyond the traditional definition of the applicability domain (AD). The practical utility of the models developed in this thesis was demonstrated by screening high-priority emerging contaminants, such as PFAS and pharmaceutical and personal care products (PPCPs), successfully filling data gaps and enabling the prioritisation of potentially hazardous compounds.

The scope of the present project was also broadened to directly support the regulatory adoption of NAMs in the EU, including the assessment of feasibility for replacing *in vivo* acute fish toxicity testing (OECD TG 203) with NAMs for aquatic hazard of neurotoxicants, thereby actively promoting the transition away from animal testing.

This thesis demonstrated that integrating scientifically rigorous, transparent, and mechanistic QSARs within the AOP framework is vital for supporting the transition toward a toxic-free environment and regulatory adoption of NAMs for TH system disruption assessment.

CHAPTER 1

General Introduction

1.1 Background

1.1.1 Chemical pollution: an environmental and human health concern

Chemical substances are integral to nearly every human activity, from agriculture and manufacturing to healthcare, making them truly essential to our modern society. Their widespread presence is a direct result of the rapid development of the chemical industry that began in the 19th century. This era brought significant scientific and industrial advancements, leading to the synthesis of a vast array of novel compounds. The sheer scale and variety of chemical substances are remarkable. For instance, over 290 million organic and inorganic substances have been catalogued with unique and unambiguous chemical abstract service registry numbers (CASRN) in the CAS REGISTRY database (www.cas.org). Meanwhile, over 350,000 chemical substances have been estimated to be registered worldwide, according to research by Wang and co-workers [1], demonstrating the vastness and complexity of the global chemical landscape. According to the report on the state of the European environment published by the European Environment Agency (EEA) in 2019, approximately 100,000 chemical substances have been estimated to be on the European market [2]. While chemical substances offer countless benefits, their widespread use carries inherent potential risks to human health and environment. Upon use and emission, exposure to chemical substances can cause adverse effects to living organisms. Therefore, even chemical substances that are beneficial to modern society can pose significant dangers if not managed adequately.

Chemical pollution has been identified as one of the nine planetary boundaries and it is considered one of the most serious threats facing the planet [3,4]. This issue is further exacerbated since it is a core component of the triple planetary crisis, a term adopted by the United Nations (UN) to describe three major and interconnected environmental challenges: climate change, pollution, and biodiversity loss (https://unfccc.int/news/what-is-the-triple-planetary-crisis). A prime example of international commitment to this issue is the 2030 Agenda for Sustainable Development, adopted by all UN Member States in 2015, outlining 17 Sustainable Development Goals (SDGs) [5]. Within this framework, reducing exposure to hazardous chemicals is thus essential for achieving multiple key SDGs, particularly SDG 3 (i.e., *good health and well-being — ensure healthy lives and promote well-being for all at all ages*), SDG 6 (i.e., *clean water and sanitation — ensure availability and sustainable management of water and sanitation for all*), and SDG 12 (i.e., *responsible consumption and production — ensure sustainable consumption and production patterns*).

Despite the widespread presence of numerous chemical substances on the market, a comprehensive understanding of their potential risks remains remarkably limited. In 2020, the EEA highlighted this significant data gap, indicating that robust hazard and

exposure evaluations is available for only a small fraction of these compounds [2]. Specifically, about 500 chemicals have undergone thorough extensive hazard and exposure evaluations, meaning that a robust set of information is available for them and their risks are relatively well understood [2]. A larger group, approximately 10,000 chemicals, has undergone a moderate level of scrutiny, with some aspects of their hazards and exposures documented and their risks characterised to a certain extent [2]. However, this situation becomes more concerning for the remaining chemical substances. Approximately 20,000 chemicals possess only limited hazard and exposure data [2]. Most alarmingly, a vast majority — around 70,000 chemicals — are very poorly understood in terms of their potential risks, meaning that they have not been adequately assessed for safe use [2]. This highlighted that, despite the widespread presence of these compounds, current understanding of their potential hazards and exposure levels varies considerably.

1.1.2 Chemical risk assessment

The safe use of a chemical substance is primarily assessed through a risk assessment process, whose results are used to inform decision-making [6]. Both environmental and human risk assessments are performed. While these two areas share key principles, their ultimate objectives differ significantly. Environmental risk assessment (ERA) aims to safeguard species populations and ecosystemic services. In contrast, human health risk assessment (HHRA) focuses on protecting every human individual [6].

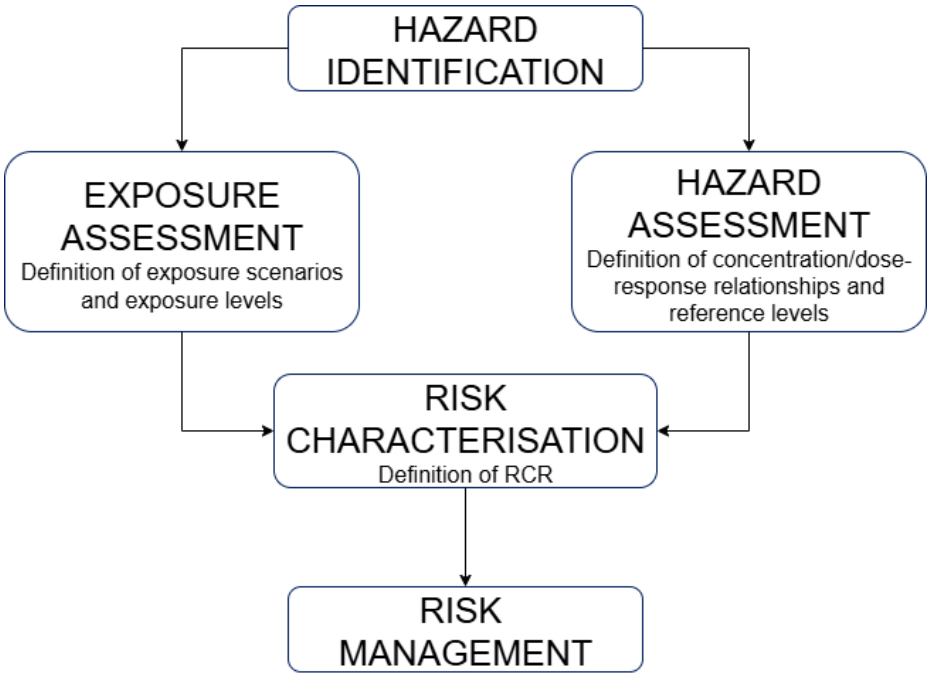

Figure 1.1. Chemical risk assessment framework.

The chemical risk assessment (CRA) process follows a structured framework, illustrated in Figure 1.1. The first step is hazard identification, where the intrinsic harmful properties of a chemical substance are defined. This is followed by hazard assessment, which determines the level of exposure at which the chemical substance is unlikely to cause harm. To accomplish this, a concentration- or dose-response relationship must first be established, allowing for the definition of a threshold value, i.e., the exposure level of the chemical substance that is not associated with any expected adverse effects. Different types of specific threshold values can be derived, varying depending on whether the risk assessment focuses on environmental or human health protection. In the context of ERA, threshold values typically include no-observed effect concentration (NOEC), lowest-observed effect concentration (LOEC), and effect concentration (EC) causing an effect in a specific percentage of the tested population, usually 10% or 50%. These threshold values are then used to derive a predicted no effect concentration (PNEC) for a specific environmental compartment, i.e., the concentration below which adverse effects on organisms are unlikely to occur. In the context of HHRA, similar threshold values are used. These include, among the others, no-observed adverse effect level (NOAEL) and lowest-observed adverse effect level (LOAEL). These values are then used to calculate a derived no effect level (DNEL), which serves as the human health counterpart to the PNEC and represents the exposure level above which humans should not be exposed.

Following hazard characterisation, the third pivotal step within CRA is the exposure assessment. This determines how living organisms come into contact with a chemical substance, and is defined by the source, magnitude, frequency, duration, and route of exposure. The aim is to build exposure scenarios and estimate exposure levels. In the

context of ERA, exposure levels are usually defined in terms of predicted environmental concentration (PEC), which is an estimated concentration of a chemical substance for a specific environmental compartment. Whenever possible, measured environmental concentration of a chemical is preferred. In the context of HHRA, exposure levels are typically expressed in terms of total daily intake (TDI), which accounts for the milligrams of chemical per kilogram of body weight per day (mg/kg body weight/day), summed up by considering exposure from all relevant routes (i.e., inhalation, ingestion, and dermal contact).

Results from hazard characterisation and exposure assessment lead to the characterisation of the risk, where the likelihood and severity of adverse effects are evaluated through the definition of a risk characterisation ratio (RCR). The threshold used is 1: a RCR value lower than 1 means safe use of a chemical, indicating the risk is adequately controlled; a RCR value greater than 1 means concern, indicating further action is needed. This latter case leads to the final step of CRA, which is risk management. Risk management typically involves implementing various types of strategies to control or reduce the identified risks. These can range from restricting the use or marketing of the chemical substance, implementing technical measures to minimise its emissions, or by ensuring personal and/or collective protective equipment.

1.1.3 Endocrine disrupting chemicals

The endocrine system is a sophisticated network of glands and organs (Figure 1.2) responsible for the proper synthesis and secretion of hormones, such as oestradiol, testosterone, thyroxine, and insulin [7]. After being synthetised and then released, hormones travel through the bloodstream to reach specific target cells located throughout the body, where they interact with specific receptors. Interactions between hormones and receptors are keys to appropriately trigger, control, and coordinate a multitude of essential physiological processes across all life cycle stages. These processes range from crucial early stages, such as embryonic development, to maintaining adult functions like metabolic regulation, reproduction, and energy balance [7,8]. The primary function of the endocrine system is to maintain hormonal homeostasis, hence ensuring that hormones production is correctly regulated and their concentrations in the blood remain stable within appropriate physiological ranges.

A robust body of the scientific literature has demonstrated that the functions of the endocrine system can be disrupted by exposure to chemical substances [7,9–11]. As introduced by the International Programme on Chemical Safety (IPCS) and the World Health Organization (WHO) "*an exogenous substance or mixture that alters the function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations*" is defined as an endocrine disruptor [8]. Over time, the term endocrine disrupting chemicals (EDCs) has been also employed to characterise the same category of substances [7,8]. Actions by EDCs can lead to a wide

spectrum of severe adverse health effects on both humans and wildlife populations [7,8,12]. For instance, impact on humans encompass thyroid dysfunctions, metabolic diseases, immunological, neurological, reproductive, and cardiovascular disorders. Impacts on wildlife include reduced fertility, reduced survival of offspring, decreased growth, behavioural alterations, which ultimately contribute to populations decline [13]. Exposure to EDCs during sensitive developmental windows (e.g., embryonic growth, early childhood, and puberty) is a critical issue. Since these life stages are highly sensitive to tiny amounts of hormones, endocrine disruption can lead to irreversible health consequences later in life, making pregnant women, foetuses and infants particularly vulnerable [14–16]. Some EDCs can cross placenta [17], while others can be transferred via lactation [7,13]. There can also be a considerable delay, or latency period, between exposure and the manifestation of adverse health effects [9]. Such effects can also be transgenerational, passed down through epigenetic alterations to future generations [18–20].

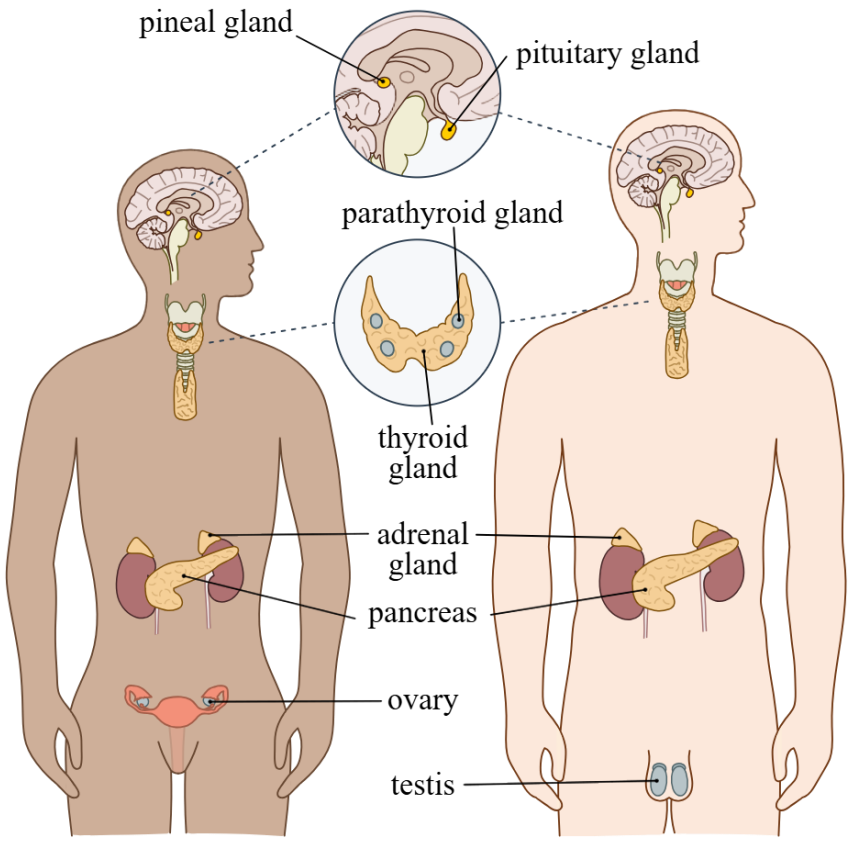

Figure 1.2. Principal endocrine glands and organs in humans. By OpenStax & Tomáš Kebert & umimeto.org, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons: https://commons.wikimedia.org/w/index.php?curid=95250832 (accessed on 26 September 2025).

Despite global concerns about endocrine-related health issues and the subsequent tightening of chemical regulations, several critical challenges remain to effectively manage the risks posed by EDCs. These challenges, already outlined over twenty years ago by the WHO [8], are still a current worldwide issue [7,21]. Unlike typical toxicological and ecotoxicological endpoints, endocrine disruption is a functional alteration, distinguishing itself as a unique kind of toxicity [8] and making its assessment and regulatory management particularly challenging. EDCs are poorly characterised. As previously described, a very small proportion of the vast number of chemicals have undergone assessment for their potential toxicity, including endocrine disruption [7]. About 1000 chemical substances have been identified as known or suspected endocrine disruptors [7], encompassing a broad spectrum of chemical classes and uses, including polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), perfluorinated compounds (PFCs) or per- and polyfluoroalkyl Substances (PFAS), pesticides (e.g., triazines, dicarboximides, atrazine), heavy metals, plasticizers (e.g., phthalates), polycyclic aromatic hydrocarbons (PAHs), bisphenols, as well as additives found in consumer products such as parabens, UV filters, and artificial musks [7,13]. Also naturally occurring compounds can be EDCs, such as phytoestrogens [22] and mycoestrogens [23]. As a consequence of their widespread use, EDCs are released into the environment through numerous pathways [24,25], leading to their pervasive contamination. Therefore, exposure to EDCs can occur through multiple routes [7,13], including ingestion of contaminated food [26], household dust [27], and drinking water [28], inhalation of contaminated air, either in the form of gases or airborne particles [29], use of consumer products (such as plastics, paints, detergents, cosmetics, and food packaging materials) [30], and occupational exposure [31]. In addition, humans and wildlife are often exposed to mixtures of EDCs from multiple sources, which might cause additive or synergistic effects. Global exposure levels vary, often being higher in urban and heavily industrialized regions. However, EDCs have been detected in remote areas due to their capacity for long-range transport and accumulation within food webs [7,13]. To further add complexity to risk assessment of EDCs, these kinds of chemical substances are characterised by strong heterogeneity in terms of chemical structures, physical-chemical properties, environmental behaviours, and mechanisms of action, which can hinder the adoption of grouping strategies and require appropriate analysis and research. Historically, scientific interest in endocrine disruption was narrowly focused on compounds interacting with nuclear hormone receptors, assuming EDCs action was primarily mediated by receptor binding [9,32]. However, recent toxicological research has significantly broadened this understanding. Nowadays, it is well established that EDCs can also act on a much broader spectrum of biological targets, interfering with hormones synthesis, secretion, distribution, metabolism, and excretion, ultimately impacting hormone delivery [9,32,33]. In this regard, an endocrine consensus statement outlining ten different key characteristics of EDCs (illustrated in Figure 1.3) has been recently defined by a panel of experts in

endocrinology and risk assessment [33], following the successful application of a similar approach for carcinogens. The aim was to provide a common ground to support hazard identification of EDCs worldwide based on mechanistic data or, when data are not sufficient, to identify relevant missing information and to highlight areas needing further research. These characteristics focused on mechanisms leading to alterations of hormones actions, and include interaction with nuclear hormone receptors, as well as alterations of synthesis, distribution and metabolism of hormones. It is sufficient for a chemical substance to exhibit one of the key characteristics to be recognised as an EDC [33].

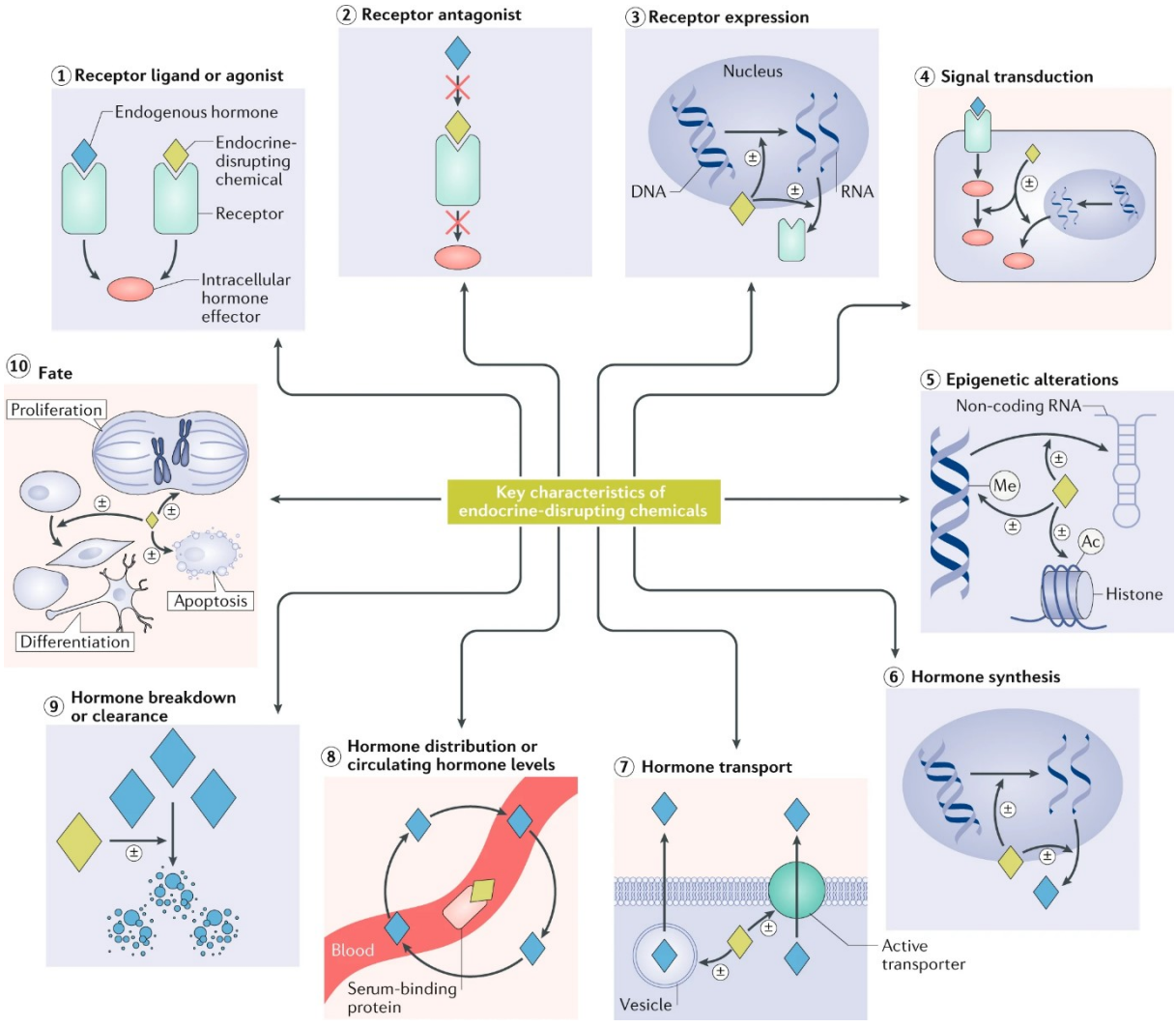

Figure 1.3. The ten "key characteristics" of EDCs. From [33], figure 1. This work is licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

1.1.4 Hypothalamic-pituitary-thyroid axis and thyroid hormone system

In vertebrates, the endocrine system relies on three major endocrine axes: the hypothalamic-pituitary-gonadal (HPG) axis, the hypothalamic-pituitary-adrenal (HPA) axis, and the hypothalamic-pituitary-thyroid (HPT) axis [8]. Among these, the HPT axis plays a key role in maintaining the overall physiological balance of thyroid hormones (THs) by controlling their homeostasis [34,35], hence it is responsible for regulating their synthesis, secretion, distribution, and metabolism. This is accomplished by operating through a sophisticated negative feedback loop, meaning that when concentrations of THs in the bloodstream become too high, the coordinated activity of the glands composing the HPT axis (i.e., hypothalamus, anterior pituitary gland, and thyroid gland) is temporarily inhibited [34,36]. This inhibition prevents excessive THs production and ensures that THs concentrations remain stable and within the proper physiological range. This intricate system is highly conserved across vertebrate species [37,38].

Proper THs homeostasis is crucial for multiple biological functions throughout the entire life cycle, from foetal development to post-natal life stages [34,35]. THs are directly involved in the development of the brain [39] and nervous system [40], as well as are essential for maintaining proper brain functions throughout life [41]. In addition, THs play a key role in regulating basal metabolism and thermogenesis [42], as well as have a key influence on cardiovascular system functions [43], reproductive regulation [44], and skeletal system development [45].

While the HPT axis is the central control mechanism for THs homeostasis, the TH system is a much broader and comprehensive network encompassing all the physiological processes and biological components involving THs throughout the body, from their synthesis to their metabolism and ultimate binding to TH receptors for inhibiting or activating gene transcription. This integrated network ensures that cells and tissues receive the appropriate amount of THs to exert their functions. The hypothalamus, located in the brain, initiates the mechanism by secreting thyrotropin releasing hormone (TRH) [42]. TRH then binds to thyrotropin releasing hormone receptor (TRHR) located in the thyrotropes at the anterior pituitary gland, stimulating the synthesis and secretion of thyroid stimulating hormone (TSH) [42]. TSH, in turn, reaches the thyroid gland, where it binds to thyroid stimulating hormone receptor (TSHR) located in the thyroid follicular cells, finally triggering the process of synthesis and secretion of THs [42]. More specifically, a sodium-iodide symporter (NIS) actively transports iodide ions (I^-) from the bloodstream into the thyroid follicular cells [46]. Then, the thyroperoxidase (or thyroid peroxidase, TPO), which is an enzyme also stimulated by TSH binding to TSHR, firstly catalyses iodide oxidation to molecular iodine (I_2) by using hydrogen peroxide (H_2O_2) as a substrate, and then facilitates the conjugation of I_2 with tyrosine residues of thyroglobulin protein [47]. The conjugation of I_2 with tyrosine residues, which serve as the precursor of THs, leads to the formation of monoiodotyrosine (MIT) and diiodotyrosine (DIT). Ultimately, TPO modulates the

coupling of MIT and DIT to generate THs: the coupling of one MIT molecule with one DIT molecule produces 3,5,3′-triiodo-L-thyronine (T3), while the coupling of two DIT molecules produces 3,5,3',5′-tetraiodo-L-thyronine (or thyroxine, T4) [47]. Finally, the secretion and release of T3 and T4 in the circulatory system undergoes by proteolysis of thyroglobulin inside the thyroid follicular cells and by mediation of TH monocarboxylate transporters [48]. Once released, T3 and T4 are transported through the bloodstream mainly bound to TH distributor proteins like thyroid binding globulin (TBG), transthyretin (TTR), and albumin, with varying binding affinities for T3 and T4 [49]. It is important to highlight that only the free circulating fraction of THs in bloodstream is biologically active and is accessible for cellular uptake [49]. Thus, among others, the primary function of these three THs distributor proteins is to maintain a stable reservoir of bound THs in the blood, effectively buffering against rapid and abnormal fluctuations in free THs concentrations [49]. Others key players of the TH system are the iodothyronine deiodinases (DIO1, DIO2, and DIO3), a group of enzymes catalysing metabolic transformations to activate or deactivate THs, typically by removing iodine atoms from their outer or inner aromatic rings [50]. Although T4 is the primary hormone secreted by the thyroid, T3 is the most biological active form [51]. Thus, a fundamental reaction catalysed by deiodinases is the removal of one iodine atom from the outer ring of T4 to generate T3. Other functions include the deactivation of THs by removing one iodine atom from the inner ring of T4 to generate 3,3′,5′-triiodo-L-thyronine (or reverse T3, rT3), or converting T3 to diiodothyronine (T2), as well as the recycling of iodine [52]. Finally, transporter proteins (e.g., monocarboxylate transporter 8, or MCT8) mediate the uptake of THs in peripheral cells, where modulation of gene transcription is led by binding of THs to thyroid nuclear receptors (TRs) α and β. T3 is the most biological active form of TH, meaning that it has a higher binding affinity to TRs than T4 [53].

1.1.5 Thyroid hormone system-disrupting chemicals

As introduced previously, the maintenance of the proper physiological balance of THs is essential. Thus, any disruption of the THs homeostasis can result in critical adverse effects. These include cognitive deficits and neurobehavioral disorders, metabolic diseases (e.g., obesity, hyperthyroidism, hypothyroidism), an increased risk of cancers, and negative impacts on the immune, cardiovascular, and reproductive systems (e.g., autoimmunity, cardiac arrhythmias, and reduced fertility, respectively) [39,54–58].

Disruption of the proper THs homeostasis may arise by the activity of a subset of EDCs defined as thyroid hormone system-disrupting chemicals (THSDCs) [59]. THSDCs can interfere with the biological targets within the TH system and modulate any of the key characteristics illustrated in Figure 1.3, ultimately leading to an imbalance in THs levels [60]. Scientific studies, to date, consistently identified a wide array of chemical substances as THSDCs, including some pesticides, perchlorate, bisphenol A (BPA) and other phenols, phthalates, UV filters, polybrominated biphenyls (PBBs), PFAS, PCBs, PBDEs, PAHs, dioxins, furans, and various metals [61–67].

A critical concern is the exposure of vulnerable populations, particularly pregnant women, foetuses, infants, and children, to THSDCs [16,68–70]. THs hold crucial importance for the proper foetal development [71,72]. Among the others, THs contribute to metabolism, differentiation, and development of the placenta [61], as well as to brain differentiation and formation of the central nervous system [73]. During the first trimester of human pregnancy, the developing foetus is entirely reliant on maternal THs supply [69]. This early period of gestation is critical as the thyroid gland of the foetus is not yet fully developed, and it does not produce adequate amount of THs until roughly mid gestation [71,74]. Hence, maternal THs, primarily T4, must be actively transported across the placenta to reach the foetus (for instance, by TTR, see Section 1.1.6). Given the crucial importance of THs for the proper foetal development, even a small deficiency of maternal THs supply during pregnancy can lead to serious and irreversible effects (e.g., neurological damages) with lasting effects in adulthood [69,71,74].

Multiple studies suggest the potential negative implications by THSDCs not only on mammals, but also on non-mammalian vertebrates such as fish [75,76], amphibians [77,78], and birds [79,80], given that the TH system is highly conserved across vertebrates [37,38]. This underscores the broad ecological concern associated with THSDCs, on the one hand highlighting their prompt identification as a critical priority, on the other hand underscoring how methods for assessing human hazard could be used for extrapolating data useful for environmental hazard assessment, and vice-versa [38].

1.1.6 Transthyretin: a TH distributor protein

In human blood, over 99% of T3 and T4 is bound to the three TH distributor proteins, i.e., TTR, TBG, and albumin, with varying binding affinities for T3 and T4 [49]. Among these, TTR stands out due to specific functional features. Human TTR (hTTR) is a homotetrameric protein with a molecular mass of 55 kDa protein composed by four identical subunits, each consisting of 127 amino acids folded into a specific beta-barrel structure (Figure 1.4) [81].

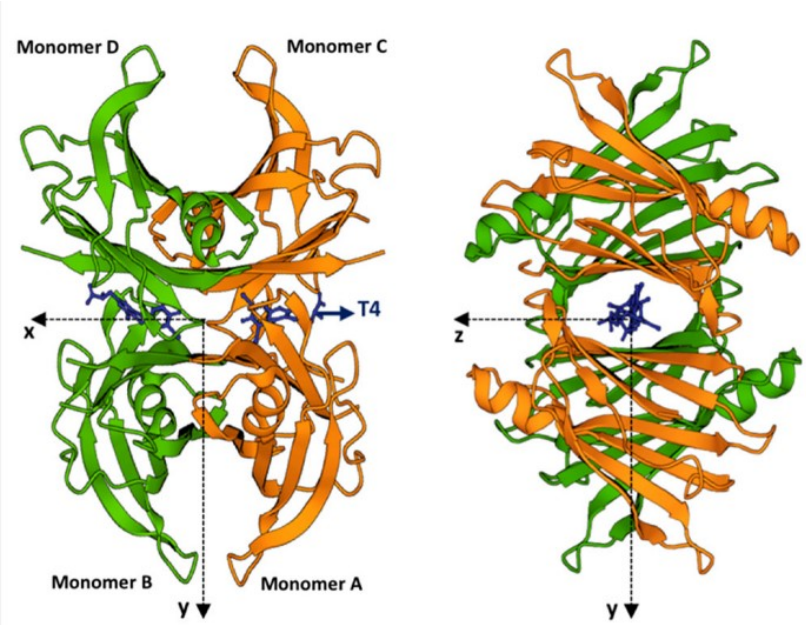

Figure 1.4. Structure of hTTR. Adapted from [82], figure 1. This work is licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/>.

TTR is remarkably similar across different vertebrate species, suggesting its crucial and conserved role in evolution [83]. Whilst TBG and albumin are mainly synthetised in the liver, hTTR is synthetised in various organs and cell types, including the liver, the choroid plexus in the brain, trophoblast, and the retinal pigment epithelium in the eye, indicating its involvement in many biological processes [84,85]. As aforementioned in Section 1.1.4, the primary function of these three proteins is to maintain a stable reservoir of bound THs in the blood and to buffer against rapid and abnormal fluctuations in free THs concentrations, so that only the free circulating fraction of THs in the bloodstream is biologically active and is accessible for cellular uptake [49]. In contrast to TBG and albumin, hTTR plays the key role in transporting T4 and retinol (vitamin A in complex with the retinol binding protein) in both the blood serum and the cerebrospinal fluid (CSF) [81]. The deliver of T4 into the CSF is absolutely essential. Unlike the others TH distributor proteins, hTTR is the only one synthetised in the choroid plexus in the brain, making it key for delivering T4 across blood-brain barrier and placenta during early foetal development. Hence, if a chemical substance binds to hTTR, this event might reduce the proper T4 supply to the developing foetus, resulting in serious and irreversible effects as anticipated previously [67]. Considering the critical functions of hTTR, it is essential to identify chemical substances capable to displace T4 from hTTR by binding to it.

In the context of the recent advent and development of the adverse outcome pathway (AOP) framework [86], TTR binding by chemicals has been recognised as a key molecular initiating event (MIE) in AOPs for TH system disruption, causing key events (KEs) such as alterations of THs concentrations, and ultimately leading to adverse

outcomes (AOs) like developmental neurotoxicity in mammalian species [38,60]. Furthermore, TTR binding has been identified as MIE in AOPs dedicated to specific adverse outcomes, like neurotoxicity [87], neurodevelopmental toxicity (see ID 152 on https://aopwiki.org/), and altered amphibians metamorphosis (see ID 366 on https://aopwiki.org/). It is worth noting that MIEs involving other types of thyroid-related biological targets, such as TR, TPO, and NIS, have been highlighted to have a direct role in leading to neurotoxicity [88–90].

1.1.7 European regulatory frameworks and future directions in advancing methods for testing and assessment

Since 1999, the European Union (EU) has consistently acted to address EDCs and other harmful chemicals to safeguard both human health and the environment [91]. The EU Green Deal is one of the most recent examples of these efforts [92]. Launched in 2019, the EU Green Deal sets, among the others, the clear objective to foster a toxic-free environment by implementing a Chemicals Strategy for Sustainability (CSS) [93]. The CSS was adopted by the European Commission (EC) in 2020 aiming to ensure that all new chemicals are intrinsically "safe and sustainable by design" and to phase out the most harmful ones, such as those affecting the endocrine, nervous or immune system, for non-essential uses [93]. This requires streamlining, harmonising, and strengthening the existing European chemical regulatory framework, achieving a horizontal EU legal approach [93]. Among more than fifty relevant pieces of legislations, particular focus is placed on key regulations such as the Registration, Evaluation, Authorisation and Restriction of Chemicals Regulation (REACH, Regulation (EC) No. 1907/2006) [94], the Plant Protection Products (PPP) Regulation (Regulation (EC) No. 1107/2009) [95], the Biocidal Products Regulation (BPR, Regulation (EC) No. 528/2012) [96], and the Classification, Labelling and Packaging Regulation (CLP, Regulation (EC) No. 1272/2008) [97]. For instance, EDCs are currently identified under REACH as substances of very high concern (SVHC), and discussions are ongoing to update this regulation by introducing standard information requirements specific for ED properties [98]. PPP Regulation and BPR were amended in 2018 [99] and 2017 [100], respectively, to include criteria to determine ED properties of active substances in plant protection or biocidal products, respectively, and to carry out specific evaluation for approval or renewal of active substances in these products. More recently, in 2022, also the CLP regulation was amended by the introduction of new hazard classes regarding endocrine disruption for human health and for the environment [101].

According to such regulations, a specific set of criteria has been established and must be met in order to identify a chemical substance as an EDC. These criteria, although slight changes in the terminology occurred (as described by Holmer and colleagues [102]), are uniformly based on the WHO/IPCS definition of endocrine disruptor [8]. The criteria include [101,103]:

1) An adverse effect;

2) An endocrine mode of action;
3) A biological causal link between the adverse effect and the endocrine mode of action.

Compared to the key characteristics mentioned previously [33], these criteria set a much higher bar for a chemical substance to be considered as an EDC, moving beyond just the potential for alteration of hormone actions.

In order to gather information and meet the ED criteria, there is still a heavy reliance on (vertebrate) animal testing [102,104,105]. This is becoming even more critical due to the introduction of new ED hazard classes to CLP and especially the upcoming revisions of REACH regulation, which is expected to increase new data requests and therefore animal testing [98,104]. Traditional toxicological and ecotoxicological hazard assessment methodologies that rely solely on (vertebrate) animal-based testing are no longer feasible, particularly following a substance-by-substance approach. These methodologies, while still widely considered the gold standard for regulatory toxicology and ecotoxicology [106], are extremely resource demanding in terms of time, economical investments, and personnel. They also present significant limitations regarding interspecies extrapolations, making it challenging to accurately predict hazards across different species based on animal-based test results [107,108]. Lastly, but of equal importance, they raise profound ethical concerns on animal welfare. This is particularly relevant given the vast number of existing chemical substances and the continuous placing on the market of novel ones [1,2]. For many of these, toxicological and ecotoxicological data are scarce or even entirely lacking [2], as described in Section 1.1.1.

To tackle these challenges, the EU is developing a roadmap aimed at heavily reducing animal testing for safety assessment of chemical substances regulated under multiple pieces of legislation [109]. To support this transition and speed up the identification of hazardous substances, the need to develop, validate, and apply new approach methodologies (NAMs) is widely acknowledged [107,110–112]. Multiple definitions of NAMs have been proposed, which consistently include *in silico*, *in vitro*, and *in chemico* approaches [107,113] and serve as alternatives to (vertebrate) animal testing following the 3Rs principles (refinement, reduction, and replacement). The push to minimise animal testing is already deeply embedded in the major EU chemical regulations like REACH, BPR, and PPP Regulation. These regulations, especially REACH, are the primary drivers for generating new hazard data on chemical substances in EU. Given the significant data requirements, these regulations consistently mandate that animal testing be used as a "last resort". For example, Article 25 of REACH explicitly states that "*in order to avoid animal testing, testing on vertebrate animals for the purposes of this Regulation shall be undertaken only as a last resort*" [94]. Exactly the same is mandated in Article 62 of the BPR [96], and a similar mandate appears in PPP Regulation [95]. This emphasis makes clear the need for data generated by means of alternatives to (vertebrate) animal testing. Under the framework of the EU's CSS [93], the development and implementation of alternative methods is heavily promoted to

reduce reliance on animal testing, even to support the identification of EDCs [98,102]. This approach is gaining urgency as the European Chemicals Agency (ECHA), in its recent "Key Areas of Regulatory Challenge" report [104], highlighted this need in order to accelerate ED assessment. Burden and co-workers specifically highlighted that assessment of ED properties based on traditional (vertebrate) animal testing is no longer feasible [98].

The AOP framework provides a crucial tool to address these needs by organising and synthetising mechanistic knowledge. An AOP is an evidence-based description of the causal chain of events, starting from one or more MIEs (i.e., the initial interaction between a chemical substance and a biological target), and progressing through a series of KEs, ultimately leading to one or more AOs at the organism or population level [86]. AOP-Wiki (https://aopwiki.org/) and AOP knowledge base (https://aopkb.oecd.org/) serve as the principal global repositories for AOPs. The design of the AOP framework is suitable for bridging gap between results from NAMs, which often operate at the molecular or cellular level, and apical endpoints. Considering how the criteria for EDC identification are structured, the integrated application combining NAMs and the AOP framework has been suggested as an effective and fit-for-purpose approach (see Figure 1.5) [114,115].

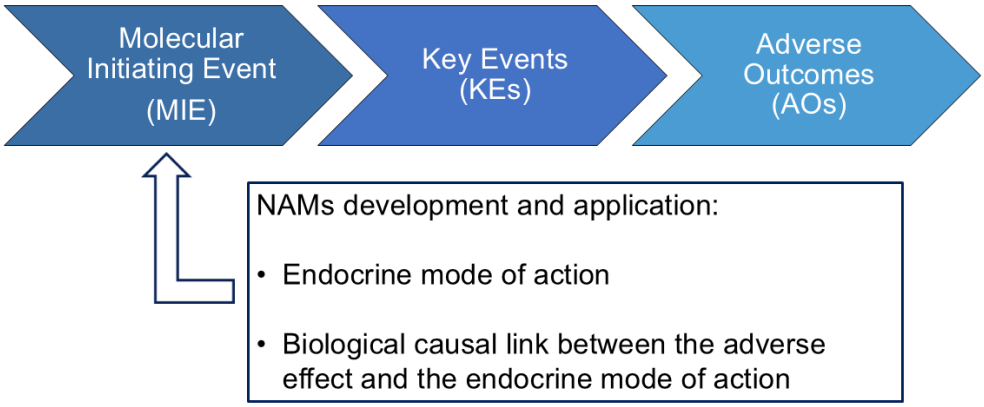

Figure 1.5. Visual representation of the NAMs-AOP integrated approach.

In order to fulfil the ED criteria, NAMs can play a key role in demonstrating an endocrine mode of action by identifying MIEs through which substances can trigger subsequent KEs and ultimately AOs, moving beyond the classical receptor-binding assessment. Furthermore, the AOP framework can contribute in defining the biologically plausible link between an endocrine mode of action and an adverse effect [102,116], as well as can provide a strong mechanistic understanding and biological relevance which are required to increase confidence in NAMs [117,118]. *In vitro* and *in*

silico methodologies have been included in the "Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals" by the Organisation for Economic Co-operation and Development (OECD) as a relevant source of information to assess ED properties of substances [119], as well as their application has also been recommended by the scientific community [120–122]. Further highlighting the need of NAMs in this context, while some existing OECD test guidelines can evaluate certain endocrine modalities, they are insufficient for comprehensively assessing all potential endocrine modes of action. This is because they often provide data only on the adverse effect, rather than the underlying mechanisms. For various chemical substances subject to regulations like REACH, BPR, and PPP Regulation there is often an adequate amount of data regarding their endocrine adverse effects, however, a major gap exists in mechanistic data availability which is essential to fulfil the criteria "endocrine mode of action". An important subset of NAMs is represented by *in silico* methods and more specifically those based on the quantitative structure-activity relationship (QSAR) approach. The synergistic use of QSARs within the AOP framework has been recommended to support the identification of EDCs, as well as of other types of toxicity, both in the scientific literature [59,121,123–126] and by the OECD [127].

In its commitment to this field, the EU has launched several key projects to develop novel testing and screening methods based on NAMs (and QSARs) for ED assessment and mode of action identification. A leading example is the European Cluster to Improve Identification of Endocrine Disruptors (https://eurion-cluster.eu/). This cluster includes a series of projects funded under the EU Horizon 2020 program, like ATHENA [59] (https://athenaedctestmethods.net), SCREENED (https://www.screened-project.eu/), ERGO [128] (https://ergo-project.eu/), OBERON [129] (https://oberon-4eu.com/), and ENDPOINTS (https://endpoints.eu/). For instance, the ATHENA project aims to generate new data and develop novel QSARs for the prediction of TH system-related MIEs [59]. Beyond EURION, the Partnership for the Assessment of Risks from Chemicals (https://www.eu-parc.eu/) is another major initiative to develop next-generation chemical risk assessment to protect human health and the environment and support the EU's CSS goals [111], with a notable focus on the development of new QSARs for MIEs predictions [113,121].

1.1.8 QSARs as NAMs

An important subset of NAMs is represented by *in silico* methods [107,113]. The quantitative structure-activity (property) relationship (QSA(P)R) approach is one of the most established *in silico* methods that aims to predict biological activities or physical and chemical properties of chemical substances based on their structural features. The key assumption underlying the QSA(P)R approach is that any variation in the activity or physicochemical property is correlated to variations in the chemical structure. To build this kind of relationship, structural features are represented in a numerical quantitative form as molecular descriptors. Hence, QSA(P)Rs are mathematical models where the dependent variable is the activity or property that is

predicted, and the independent variables are the molecular descriptors encoding for the structural features of chemical substances. Thousands of molecular descriptors are used [130], spanning from simple and experimental physicochemical properties (e.g., hydrophobicity, molecular weight) to more sophisticated theoretical descriptors (e.g., topological indices). Multiple statistical or machine learning algorithms are employed to build a QSA(P)R model, such as multiple linear regression (MLR), k-nearest neighbour (kNN), support vector machines (SVM), and neural networks (NN) [131]. The choice of the algorithm can be influenced by several factors, including the complexity of the training data (e.g., linear or non-linear relationships between dependent and independent variables), the desired interpretability of the model, and the field of application. Indeed, QSA(P)R approaches have wide-ranging applications in drug discovery and drug design, toxicology and ecotoxicology, and environmental chemistry [123]. In drug discovery and drug design, QSA(P)R models are used to predict activity of new drug candidates, to identify structural requirements to optimise selectivity and efficacy, as well as to estimate pharmacokinetic properties. In toxicology and ecotoxicology, the QSA(P)R approach is used to estimate potential hazards posed by chemical substances to human health (e.g., skin sensitisation) or ecosystems (e.g., aquatic toxicity), ultimately to support CRA. In environmental chemistry, these approaches are employed to predict physical-chemical properties, such as water solubility and volatility, ultimately to estimate transport and distribution, and therefore the fate, of chemicals across environmental compartments.

Since these computational tools gained importance in chemical regulations, the need for a global, standardised approach became clear. To ensure predictions from QSA(P)R models are scientifically sound and trustworthy, and to promote their acceptance across different countries, the OECD established a set of five key principles to provide a conceptual framework for developing, validating and reporting QSA(P)R models, and ultimately to ensure their reliability and scientific integrity [132]. The five OECD principles are:

1) *A defined endpoint;*
2) *An unambiguous algorithm;*
3) *A defined applicability domain;*
4) *Appropriate measures of goodness-of-fit, robustness, and predictivity;*
5) *A mechanistic interpretation, if possible.*

Principle one demands that every QSA(P)R model must predict a clearly defined endpoint, i.e., any biological activity or physicochemical property. This is crucial for transparency, as the same activity or property can be measured using different experimental methods, and therefore can be expressed in a different way. Ideally, the data used to train the model should be as consistent as possible, and derived from a single experimental protocol. Principle two underscores the need for transparency and reproducibility of the mathematical process used to generate predictions. This principle requires a detailed description of the data used, the molecular descriptors, the modelling procedure, and the statistical methods. Without this clarity, it is not

possible for an independent party to verify and reproduce the results from a model. Principle three addresses the need to define the applicability domain (AD) of a QSA(P)R model, hence to define the space where the model is expected to provide reliable predictions for new chemical substances, since a single QSA(P)R model cannot reliably predict outcomes for every molecular structure. This space is determined by the characteristics of the data used to build the model, thus by the structural, physicochemical, and experimental response information of the training set. According to principle four, a valid QSA(P)R model must be supported by strong statistical evidence of its performance. This involves evaluating its internal robustness and external predictivity when it is applied to unseen molecular structures. Finally, principle five encourages, but does not strictly require, the definition of a plausible mechanistic link between the molecular descriptors selected in the model and the predicted response. The presence of a mechanistic interpretation can add confidence and a deeper scientific understanding to the predictions from the model, thereby enhancing the model's credibility.

The OECD principles provide the scientific foundation for validating QSA(P)R models. However, simply following them is not enough. The information must be communicated clearly in order to make models and their predictions accepted. To this end, specific reporting formats were designed: the QSAR Model Reporting Format (QMRF) and the QSAR Prediction Reporting Format (QPRF). The QMRF includes all the relevant information about the model, such as the training set, the modelling approach, and the definition of the AD. The QPRF includes all the relevant information about the chemical substance under analysis, about the applied model, and about the prediction (e.g., if the prediction falls inside the AD). To further enhance the evaluation of the reliability of QSA(P)R predictions and results, and their acceptance, for regulatory purposes, the QSAR Assessment Framework (QAF) was published by the OECD in 2023 [133,134]. This new guidance, which is built upon the five OECD principles, establishes new checklists and assessment elements to be used as strict criteria for the regulatory assessment of models, predictions, and results based on multiple predictions.

In addition to the OECD principles, a set of "good practices" have been proposed in the scientific literature for the proper development and validation of QSA(P)R models [135,136]. The modelling workflow adopted in the present thesis is illustrated in Figure 1.6.

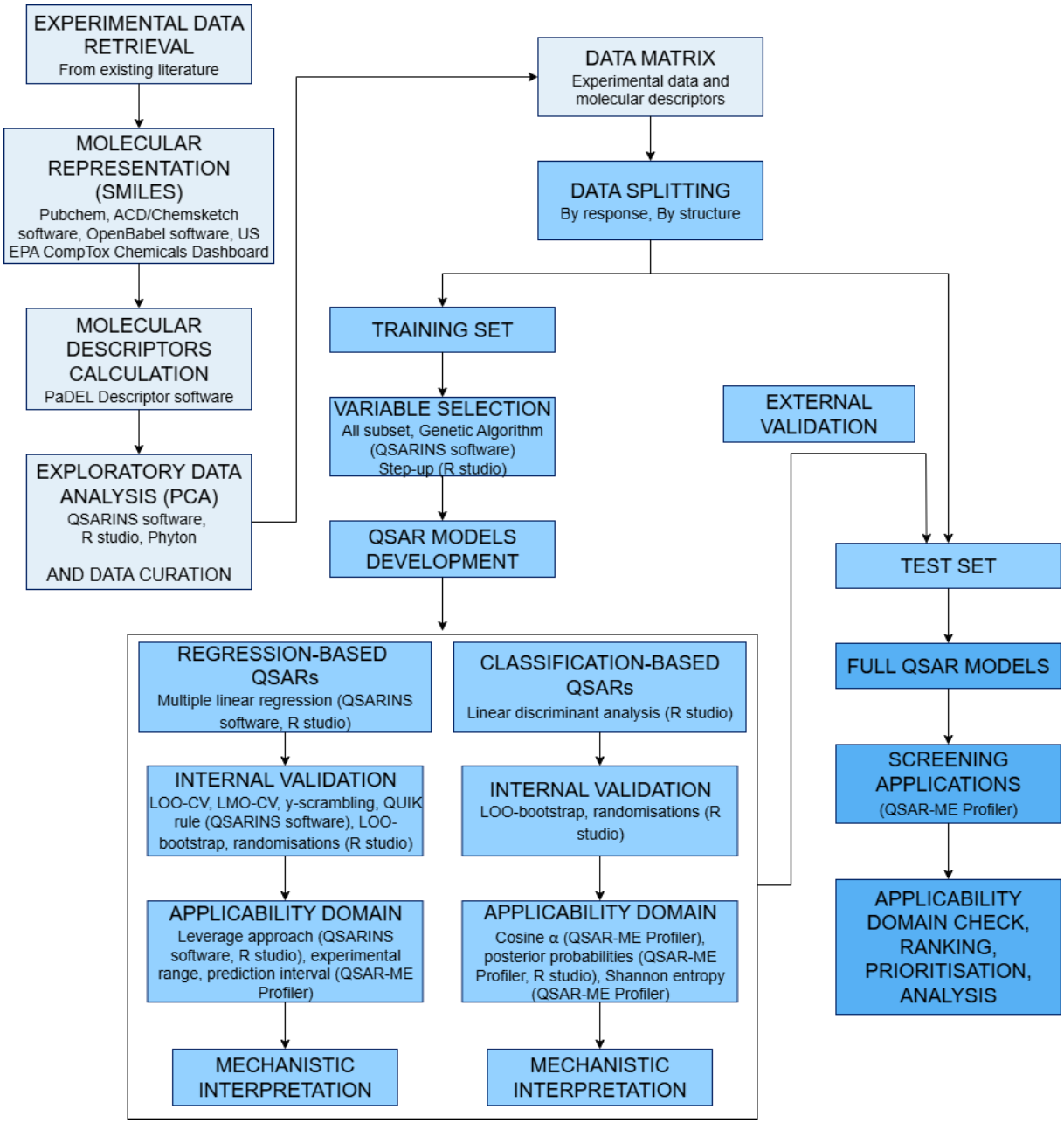

Figure 1.6. Modelling workflow followed in the present thesis.

1.2 Methods

1.2.1 SMILES and molecular descriptors

Molecular structures of chemical substances must be represented in a format that is readable and processable by computational tools, such as the simplified molecular entry line system (SMILES) notation. In particular, SMILES defines a linear notation that allows for the representation of the molecular structure of a chemical substance as a simple text string [137]. The encoding system of the SMILES notation approach is bult upon a set of fundamental rules establishing how to represent and specify the presence of atoms, bonds, branching, aromaticity, molecular topology, and so on.

In this thesis, SMILES notations were used as input to compute theoretical molecular descriptors, which numerically quantify structural and topological features of a molecular structure. Currently, thousands of molecular descriptors have been defined and can be used for QSAR modelling purposes [130]. They can be organised in the following four main categories, according to the level of complexity they encode for:

- 0-dimensional (0D) descriptors are the simplest ones, given that they can be extracted directly from the molecular formula of a chemical substance without taking into account any information about, for instance, atom connectivity. Typical examples of 0D descriptors include molecular weight and atom counts.
- 1-dimensional (1D) descriptors encode for the presence (or absence) and the frequency of substructures and fragments. Typical examples of 1D descriptors include the presence (or absence) of functional groups.
- 2-dimensional (2D) descriptors add a layer of complexity, since they take into account the adjacency and connectivity of atoms in a bidimensional space. For this reason, they are often defined as topological descriptors. Typical examples of 2D descriptors include autocorrelation descriptors and topological indices.
- 3-dimensional (3D) descriptors encode for the spatial configuration of the entire molecular structure, and therefore how atoms are organised in a three-dimensional space. For this reason, they are often defined as geometrical descriptors. The computation of 3D descriptors requires a preliminary geometric optimisation of the molecular structure, and they mainly find practical application in drug discovery and pharmacological research. 3D descriptors were not used in this thesis.
- Fingerprints (FPs) are binary vectors encoding for the presence or absence of fragments within a molecular structure. Thus, FPs offer a comprehensive way to encode all the structural components (i.e., fragments) of a molecular structure as a string of ones and zeros. A wide range of different algorithms can be applied for their computation, depending on how the fragments are defined.

1.2.2 Principal component analysis

Principal component analysis (PCA) [138] is the most known projection technique employed for dimensionality reduction when dealing with multivariate datasets, allowing for exploring correlations among variables (e.g., molecular descriptors), relationships among objects (e.g., chemical substances), and for synthetising data by eliminating noise and irrelevant information. At its core, PCA is a process of rotating the original variables included in a multivariate dataset by linearly combining them to create a set of new variables called principal components, aiming to reproduce the variance/covariance structure of the original variables. The rotation is performed so that the first principal component is oriented in the direction of maximum data variance. The second principal component is perpendicular to the first, and it is oriented with the next highest variance, and so on for all subsequent components. The number of principal components is the same as the number of original variables. An important characteristic of the principal components is that they are orthogonal. PCA allows for condensing most of the information contained in the original variables in the first principal components, making the analysis more manageable. Hence, a restricted number of significant principal components accounts for the most of the explained variance. The mathematical procedure for PCA results in two important matrices: the loading matrix and the score matrix. The elements of the loading matrix, called loadings, represent the weight of each original variable on the principal components. The loading matrix is a rotation matrix that, when multiplied by the original multivariate dataset, transforms it into a new orthogonal space. The resulting matrix from this transformation is the score matrix. The elements of the score matrix, called scores, are the new coordinates of the objects with respect to the principal components. Ultimately, the data included in the loading and score matrices can be used for graphical visualisation of the relationships among the original variables (loading plot) and the objects (score plot). The loading plot is created using data from the loading matrix, and its utility lies in studying the weight of each original variable and their potential correlations with respect to a specific principal component. The score plot is generated using data from the score matrix, and it allows for determining the relative position of the objects in the new orthogonal space. Its utility lies in identifying the presence of patterns, groups, and outliers. For a comprehensive view, a biplot can be generated by integrating the loading plot and the score plot, i.e., by drawing both the scores and the loadings.

1.2.3 Modelling methods

A breadth of algorithms can be employed on a multivariate dataset for modelling purposes [131]. In this thesis, multiple linear regression (MLR) by means of the ordinary least squares (OLS) was used as a method to develop regression models (see Chapter 2 and Chapter 4), while linear discriminant analysis (LDA) was employed to develop classification models (see Chapter 4).

In the case of MLR, a relationship is sought between a dependent variable (i.e., a vector containing the experimental responses, y) and a matrix of independent variables (i.e., the molecular descriptors, X), expressed as a vector of coefficients (b) [139]. A model is defined as linear if the experimental response is a linear combination of the variables in the model. A MLR model calculated over p variables can be expressed according to the following algebraic formula:

$$y_i = b0 + b1\, x_i1 + b2x_i2 + \cdots \ldots . bpx_ip \quad (1.1)$$

where y_i is the predicted response of the *i-th* object, $b0$ is the intercept, $b1..bp$ are the coefficients, and $x_i1..x_ip$ are the p independent variables of the *i-th* object.

The OLS method allows for the identification of the best regression capable of minimising the response error. This error can be expressed in terms of residual sum of squares (RSS) between the observed responses and the responses predicted by the model for all of the objects.

In the case of LDA, a relationship is sought between a dependent variable (i.e., a vector containing the experimental classes) and a matrix of independent variables (i.e., the molecular descriptors) through a set of discriminant functions [139]. The aim is to identify linear combinations of the independent variables that best separate the different classes of the dependent variable. In other words, the core principle is to maximise the ratio of between-class variance to within-class variance. The most appropriate discriminant functions, i.e., those that allows for minimising the misclassification rate, are identified by this method to effectively classify new, unknown objects. To classify a new object, the posterior probability of that object belonging to each possible class is calculated. Thus, the object is assigned to the class with the highest posterior probability.

1.2.4 Variable selection

The variable selection seeks to identify the best combination of molecular descriptors of a QSAR model. When the biological mechanism under scrutiny is known a priori, it is possible to select the molecular descriptors based on this knowledge. When the phenomenon is not adequately known, different types of methodologies can be applied to select the best set of molecular descriptors, starting from a large pool.

In this thesis, three types of variable selection approaches were employed: the all-subset selection (see Chapter 2), the genetic algorithms (see Chapter 2), and the step-up procedure (see Chapter 4).

- The all-subset selection method evaluates all the possible combinations of molecular descriptors from a pool in order to identify the best-performing subset. Hence, it starts by developing and assessing all single-descriptor models, then progresses to consider all two-descriptors models, and continuing this process until every single possible combination of descriptors up to an arbitrary number has been developed and evaluated. Although exhaustive, it is computationally intensive. In this work, it was employed to explore combinations of up to two molecular descriptors.
- The genetic algorithms (GAs) [140] are methodologies inspired by the principles of biological evolution and, therefore, the terminology adopted also reflects this. The process starts with a random initialisation of a population of chromosomes (i.e., models). Each chromosome is a binary vector where each position, or gene, corresponds to a specific variable. A gene value of 1 indicates that the variable is included in the model, while 0 means it is not. This initial population is then evaluated using a fitness function that quantifies the performance of the combination of variables in each chromosome. Then, an iterative cycle of generating and refining the population is initialised, which is performed through two main operations: crossover and mutation. Crossover involves selecting, from the current population, two parent chromosomes which "mate" by exchanging portions of their genetic material (i.e., genes) to produce new offspring chromosomes. Following crossover, the new chromosomes undergo mutation. This is a random process where a gene value can be flipped from 0 to 1 or vice versa, thereby introducing new variable combinations that may not have existed in the parental population. The purpose of mutation is to prevent the algorithm from getting stuck in local optima and to increase population diversity. After the new chromosomes are generated, their fitness is evaluated and compared with the parental chromosomes. Throughout the whole procedure, the best-performing individuals, whether they are new offspring or existing parents, are kept for the next generation, while the worst-performing ones are discarded. This process is repeated over many generations until a stopping criterion is met, such as reaching a maximum number of iterations (method used in this thesis), or when the overall fitness of the population stabilises.
- The step-up procedure [141] is a kind of forward selection that starts by building a separate model for every single descriptor in the entire pool. Every single-descriptor model is then evaluated for its predictive ability using a

performance metric of choice. From this initial set, only the best-performing models are selected to move on to the next step, where one descriptor is cyclically extracted from the whole pool and added to all models. Therefore, two-descriptors models are generated and their performances are evaluated. Only the best-performing models are selected and allowed to continue the process up to the desired number of model variables. As for GAs, the final result is a population of best-performing models.

1.2.5 Validation of the models

Since the primary objective of a QSAR model is to accurately predict a response for new chemical substances, a critical step is its validation [132]. The validation process allows for the evaluation of the robustness and predictive ability of QSAR models, and typically includes two methodologies: internal validation and external validation. Internal validation tests robustness and internal stability of a model. External validation evaluates the predictive ability of a model by applying it to an independent set of data for which the response values are known. The fourth OECD principle for validating QSAR models is founded upon the importance of validation [132].

Regarding internal validation, the following techniques were employed in this thesis:

- Cross-validation (CV) technique is the most popular methodology to perform internal validation [132]. Generally, the most common form of CV is the *k*-fold CV, where the leave-one-out (LOO) CV is a particular case. The *k*-fold CV splits the training set into a *k* number of equal-sized partitions, known as "folds". For each round, one fold is held back as the validation set, while the remaining folds are used to train a model. This process is repeated until every fold has been used as validation set exactly once. The final performance is the average of the results from all rounds. The LOO CV is the simplest form of CV, where *k* is set equal to the number of objects. A single object is removed from the training set, and a model is developed using the remaining ones. The model is then applied to predict the response for the remaining object. This is repeated for every object, ensuring each one is left out once. The leave-more-out (LMO) CV is a variation of the *k*-fold CV where, cyclically (the number of cycles is determined by the user), a predetermined number of objects is removed as a validation set, and then re-entered before the next cycle.
- The QUIK rule was applied to estimate multicollinearity of the selected models (see Chapter 2) [142]. Multicollinearity is a condition where the molecular descriptors of a model are highly correlated with each other. The aim of the QUIK rule is to ensure that the correlation of the molecular descriptors with the

modelled response is greater than the correlation among the molecular descriptors. This is assessed using the K correlation index, which measures the overall correlation within a set of variables. First, the overall correlation between the block of modelling molecular descriptors and the response is calculated (K_{xy}). Then, the overall correlation among the modelling molecular descriptors is calculated (K_{xx}). The condition $K_{xy} > K_{xx}$ must be satisfied to exclude that the presence of multicollinearity is negatively affecting the quality of the model.

- The bootstrap method is a resampling technique employed to detect and mitigate overfitting in the variable selection procedure (see Chapter 4) [139]. The process first involves the creation of multiple bootstrapped training sets from the original training set by sampling with replacement (meaning that some objects in the bootstrapped training set may be duplicated, while others are left out). Then, for each new bootstrapped training set, a model is developed and it is applied on the objects that were left out. The prediction errors of the left-out objects are collected and used as an estimate of the performance of the model under scrutiny (i.e., the one developed from the original training set). This process is repeated multiple times aiming to a thorough and representative evaluation.
- The Y-scrambling methodology is a permutation test used to check for chance correlation of the selected models, i.e., to estimate the coincidental relationship between the selected modelling molecular descriptors and the response (see Chapter 2 and Chapter 4). The methodology first randomly shuffles the response values in the original dataset. In this way, any real structure-activity relationship between the modelling descriptors and the response should break. Then, a new model is developed using the same modelling molecular descriptors but with the scrambled response values, and the performance of the new scrambled model is evaluated. This process is repeated numerous times, thus generating multiple scrambled models whose individual performances are subsequently averaged. Finally, the performance of the original model (i.e., those developed with the unscrambled response) is compared with the average performance of the scrambled models. To consider a model valid and not affected by chance correlation, the average performance of the scrambled models must be much lower.
- To estimate the probability of coincidental relationship between the molecular descriptors and the modelled response because of the variable selection procedure, a randomisation technique was employed (see Chapter 4) among those proposed by Rücker and colleagues [141]. The selected method involves shuffling the descriptors values while leaving the response values unchanged. In this thesis, values of the molecular descriptors were randomised by constraining them to the range of values and/or nature of each descriptor. This differs from Rücker and colleagues [141], who did not apply such constrain to the variables. This variation from the original technique is because the complete

shuffling of the descriptors values without considering their nature could lead to more optimistic results. This can represent an issue for descriptors like FPs, which can only assume values of 0 or 1. Then, the variable selection procedure was performed using randomised descriptors and repeated multiple times to develop populations of models, and their performance was evaluated. If the performance of the model developed with the original dataset (i.e., with un-randomised descriptors) is only slightly better than the average performance of the models developed with randomised descriptors, it suggests that the model may be affected by chance correlation.

The parameters used to assess the results from internal validation will be examined on a case-specific basis throughout this thesis.

Regarding the external validation, the ideal scenario is the availability of a completely new and independent set of chemical substances for which the endpoint values are known. However, because new data can be hard to find, a more common strategy is to split the initial dataset into two subsets, i.e., the training set and the test set. The training set is used to develop the model, while the test set is used solely for the external validation. Several partitioning strategies are employed to split the initial dataset, where the main challenge is to ensure that both sets are adequately representative of the original data. In this thesis, the following two types of dataset partitioning schemes were employed: *by response* (see Chapter 2), and *by structure* (see Chapter 4) [143].

- The *by response* approach can be used when dealing with quantitative endpoints. In this case, the compounds are first sorted by the value of the response. Then, a common practice is to select one out of every two or three compounds and then assign them to the test set. All the remaining compounds are assigned to the training set. This approach helps ensure that both the training and the test set are representative of the original data in the range of the experimental responses.
- The *by structure* approach can be used irrespective of the response type, since it is not included in the process. First, PCA is performed on the original dataset. Then, the chemicals are ranked based on their scores along the first principal component. Similarly to the *by response* approach, this ranking is then used to select the portion of chemicals that is assigned to the training set or to the test set. This approach helps ensure that both the training and the test set are representative of the original data in terms of chemical structures.

The parameters used to assess results from external validation will be examined on a case-specific basis throughout this thesis.

1.2.6 Applicability domain

The definition of the AD is a critical step in QSAR modelling studies, and its paramount importance is highlighted in the third OECD principle for the validation of QSARs [132]. The AD defines the space of a QSAR model where it is expected to provide reliable predictions for new chemical substances [144]. This space is determined by the characteristics of the data used to build the model, thus by the structural, physicochemical, and experimental response information of the training set. Thus, predictions are only considered reliable if the chemical substance being predicted falls within this defined AD. Any prediction made for a chemical substance that lies outside the AD is considered an extrapolation, and therefore is likely being unreliable.

There is not a single and universally accepted methodology for AD definition. Various approaches are currently available and employed [132,145], each with its own level of restrictiveness. In this thesis, the following approaches were used to define the ADs:

- The leverage approach was used to define the AD of the regression QSARs developed throughout this thesis. It is a commonly used methodology to quantify the distance of each chemical substance from the centroid of the chemical space of the training set. First, the HAT matrix (or leverage matrix) is calculated as

$$HAT = X(X^T X)^{-1} X^T \tag{1.2}$$

  where X is the matrix containing the molecular descriptors of the model. Each diagonal element of the HAT matrix defines the leverage value for a given chemical substance. A low leverage value indicates that a chemical substance is close to the centroid of the space defined by the training set, therefore it is similar to most of the training set chemicals. Conversely, a high leverage value means that the chemical substance is far from the centroid, thus it can be considered structurally different and, therefore, it may be judged as an outlier. In addition, a high leverage value may indicate the relevant influence of the chemical substance on the variables of the model. For new chemicals, a high leverage value may indicate that the predicted response is an extrapolation and may be unreliable. The boundaries of the AD are defined by a specific cut-off value, often represented as h^*. This threshold is defined as $3(p+1)/n$, where p is the number of descriptors in the model and n is the number of chemical substances in the training set. A chemical substance with a leverage value greater than h^* is considered to be outside the AD. Furthermore, predictions for new chemicals were considered unreliable if they fell outside the experimental response range of the training set.

- In this thesis, the AD of the classification QSARs was defined by combining both a chemical structural domain and a response domain. The chemical structural domain was defined using a distance-based approach. In particular, the distance of each chemical structure was determined by calculating the average of its three nearest neighbours based on the cosine α metric [146,147]. A compound was deemed as an outlier if this distance was smaller than the 0.95 quantile of the k-nearest neighbour within the training set's distance distribution. The response domain was defined in terms of the posterior probabilities of the classification event [146,147].

An additional layer of reliability evaluation was introduced by quantifying the uncertainty associated with each model and prediction. Individual predictions were deemed unreliable if their uncertainty was greater than the greatest uncertainty in the training set.

- For predictions from regression QSARs, the uncertainty was defined in terms of prediction interval as follows [148]:

$$prediction\ interval = \pm t_{stud\ \alpha/2} \times s \times \sqrt{1 + HAT_i} \quad (1.3)$$

Where $t_{stud\ \alpha/2}$ is the t-student calculated for $\alpha/2$ = 0.025 and s is defined as $\sqrt{\frac{\sum_n (y_i - \hat{y}_i)^2}{n-p-1}}$, where n is the number of objects in the training set, p is the number of molecular descriptors in the model, y_i is the experimental value of the i-th object, and $\hat{y}_i$ is the predicted value of the i-th object.

- For predictions from classification QSARs, the uncertainty was defined in terms of Shannon entropy as follows:

$$Shannon\ entropy = -\sum_{i=1}^{n} p_i \log p_i \quad (1.4)$$

Where p_i is the posterior probability of the event i.

1.2.7 Software

QSARINS software [143] was central to the modelling workflow employed in the work presented in Chapter 2 (paper I in list of publications). First, it was used to filter the initial set of molecular descriptors, effectively reducing redundant information.

Second, the software employed the all-subset method and a GA for variable subset selection procedure to develop a population of QSAR models. Finally, the software calculated key statistical metrics and generated diagnostic plots, which were used to evaluate the performances of the QSAR models. In the same study, ACD/ChemSketch software [149] version 2021.1.1 was used to draw the chemical structures of the compounds and to generate their corresponding SMILES notations. SMILES were canonicalized using Open Babel software [150] version 2.4.1 in both studies presented in Chapter 2 and Chapter 4 (paper II). PaDEL Descriptor software [151] version 2.21 was used to calculate theoretical molecular descriptors and fingerprints, from SMILES, in both studies presented in Chapter 2 and Chapter 4. R version 4.3.2 within RStudio version 2023.12.1+402 (https://posit.co/, last accessed 08 October 2025) was used to compute PCA and to generate plots and figures included in paper I. QSAR-ME Profiler software beta version 1.02 (available at https://dunant.dista.uninsubria.it/qsar/) was employed for the application of the QSAR models presented in Chapter 2 and in Chapter 4. R version 4.3.2 within RStudio version 2023.12.1+402 was central to the modelling workflow employed in the work presented in Chapter 4. R studio was used to filter the initial set of molecular descriptors, to perform variable subset selection, to calculate key statistical metrics, to compute bootstrapping cross-validation and randomizations, to identify outliers, to calculate distance and similarity matrices. Python version 3.10 was used to generate plots and figures included in Chapter 4 and Chapter 5 (paper III). At ECHA, IT tools and software like IUCLID, REACH-IT, R4BP3, and Dynamic Case were employed to retrieve and collect aquatic toxicity data of neurotoxic compounds. Derek Nexus software version 6.3.0 (https://www.lhasalimited.org/)) was employed to predict neurotoxicity based on structural alerts.

1.3 Aims of the PhD project

The aim of the present PhD project was to address the current critical need for a more efficient and ethical approach to chemical safety assessment. Following the priorities set out by the EU' CSS towards a toxic free environment and the need to reduce reliance on (vertebrate) animal-based testing [93], this project aimed to advance the development and regulatory uptake of NAMs to assess highly harmful chemicals, such as THSDCs.

The TH system, an important component of the endocrine system, has been historically understudied compared to other endocrine modalities. As current research is expanding the understanding of endocrine disruptors beyond nuclear receptor binding to include a broader spectrum of disruption mechanisms, the development of new QSAR models and their synergistic integration with the AOP framework is recommended to provide a deeper, mechanistic understanding and facilitate their regulatory uptake. In this project, the main focus was posed on the development of new QSAR models for predicting the ability of chemical substances to bind to hTTR, a key MIE in AOPs for TH system disruption. In addition, case studies were proposed by applying the new QSAR models to chemical substances of emerging concern, such as PFAS and pharmaceutical and personal care products (PPCPs), demonstrating how QSAR models can be applied to support hazard assessment of chemicals by filling data gaps. Ultimately, the new QSARs were implemented in the non-commercial software QSAR-ME Profiler beta version 1.02 (available free of charge at https://dunant.dista.uninsubria.it/qsar/) to facilitate their use, enabling clear assessment of ADs, of prediction uncertainties, and structural similarity analysis.

An additional challenge in the uptake of NAMs is the specific difficulty in assessing compounds that may lead to neurotoxicity. While progress has been made in assessing the feasibility to replace the acute fish toxicity study (OECD TG 203) with NAMs for environmental hazard assessments, key uncertainties remain with these chemicals. This highlights a critical need for further work on a synergistic approach between different NAMs, including QSARs, to facilitate the transition away from animal testing in the EU chemical regulatory framework.

The following objectives were pursued to achieve these aims:

1) Development and application of new QSAR models for predicting hTTR binding. This central objective of the work was achieved through a series of focused studies, presented in Chapter 2, Chapter 3, and Chapter 4:
    - The first study (Chapter 2, paper I in list of publications) was dedicated to the creation of a new dataset of hTTR binding for heterogenous compounds. This dataset was then used to create three *in vitro* assay-specific sub-datasets each including data measured with the same *in vitro* methodology. Each of these three sub-datasets was used to develop an *in vitro* assay-specific QSAR model. A heavy focus was placed on the mechanistic interpretation of the molecular descriptors to investigate the

most relevant structural features and/or properties contributing to hTTR binding.

- In a separate work, the new QSAR models presented in Chapter 2 were applied to screen potential TH system disruption by PPCPs, a class of emerging contaminants (see Chapter 3, paper IV in list of publications).
- A third, distinct study (Chapter 4, paper II in list of publications), was first focused on the development of new QSAR models tailored to predict hTTR binding by PFAS, another class of contaminants of emerging concern. Particular emphasis was posed on the mechanistic interpretation of molecular descriptors in order to elucidate the key structural features and properties driving hTTR binding by PFAS. Following their development, the new QSARs were applied to screen the most recent OECD List of PFAS in order to identify and highlight potential TH system disruptors.

2) Definition of the state-of-the-art of QSAR models for TH system disruption, which is covered in Chapter 5 (paper III in list of publications). A foundational objective was to perform a comprehensive review of existing QSAR models published in the scientific literature to predict relevant MIEs for potential TH system disruption. This analysis was conducted to provide an overview of the field by examining key aspects such as modelled endpoints, chemical datasets, modelling approaches, and relevant molecular descriptors, thereby providing a solid foundation for future research priorities.

3) Support the regulatory adoption of NAMs in the EU. This project extended beyond model development for potential TH system disruption to include a direct contribution to regulatory science. The scope of this work, presented in Chapter 6, was intentionally broadened to encompass neurotoxicity, which is known that can be led by TH system disruption. This was achieved by collaborating with ECHA to assess the feasibility of replacing *in vivo* acute fish toxicity testing (OECD TG 203) with NAMs for aquatic hazard, starting from a reference list of neurotoxic substances. This list was purposely generated for NAMs validation. Since this work was conducted under confidentiality agreement, data involving third parties are not shown in this thesis.

- The first study was dedicated to generating a new dataset of aquatic toxicity data of neurotoxic substances included in the reference list. The aim was to perform a sensitivity analysis across the three aquatic trophic levels (i.e., fish, algae, and aquatic invertebrates) in order to evaluate whether the replacement of OECD TG 203 with alternatives is feasible for this type of chemicals.
- The second study involved searching for QSAR models and tools designed to predict neurotoxicity. All the models were accurately evaluated for their reliability and applicability, and ultimately their actual predictive performances were assessed by applying them towards the neurotoxic substances included in the reference list. The aim was to evaluate whether existing and applicable QSAR models can be reliably integrated under a weight of evidence (WoE) approach along with other NAMs.

1.4 References

[1] Wang, Z., Walker, G.W., Muir, D.C.G., Nagatani-Yoshida, K., 2020. Environ. Sci. Technol. 54, 2575–2584, https://doi.org/10.1021/acs.est.9b06379.

[2] European Environment Agency, 2020. The European environment – state and outlook 2020, https://doi.org/10.2800/96749.

[3] Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., Sörlin, S., 2015. Science 347, 1259855, https://doi.org/10.1126/science.1259855.

[4] Persson, L., Carney Almroth, B.M., Collins, C.D., Cornell, S., de Wit, C.A., Diamond, M.L., Fantke, P., Hassellöv, M., MacLeod, M., Ryberg, M.W., Søgaard Jørgensen, P., Villarrubia-Gómez, P., Wang, Z., Hauschild, M.Z., 2022. Environ. Sci. Technol. 56, 1510–1521, https://doi.org/10.1021/acs.est.1c04158.

[5] United Nations, 2015. Transforming our world: The 2030 Agenda for Sustainable Development, https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf (last accessed on 06 October 2025).

[6] Leeuwen, C.J. van, Vermeire, T.G., 2007. Risk Assessment of Chemicals: An Introduction. Second Edition, Springer.

[7] Gore, A.C., La Merrill, M.A., Patisaul, H.B., Sargis, R., 2024. Endocrine Disrupting Chemicals: Threats to Human Health. The Endocrine Society and IPEN. ISBN # 978-1-955400-23-7, https://www.ipen.org/documents/endocrine-disrupting-chemicals-threats-human-health.

[8] World Health Organisation (WHO), 2002. Damstra, T., Barlow, S., Bergman, A., Kavlock, R., Van Der Kraak, G., editors. Global assessment of the state-of-the-science of endocrine disruptors. Geneva: World Health Organization, https://www.who.int/publications/i/item/WHO-PSC-EDC-02.2 (last accessed on 06 October 2025).

[9] Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocr. Rev. 30, 293–342, https://doi.org/10.1210/er.2009-0002.

[10] Zoeller, R.T., Brown, T.R., Doan, L.L., Gore, A.C., Skakkebaek, N.E., Soto, A.M., Woodruff, T.J., Vom Saal, F.S., 2012. Endocrinology 153, 4097–4110, https://doi.org/10.1210/en.2012-1422.

[11] Gore, A.C., Chappell, V.A., Fenton, S.E., Flaws, J.A., Nadal, A., Prins, G.S., Toppari, J., Zoeller, R.T., 2015. Endocr. Rev. 36, E1–E150, https://doi.org/10.1210/er.2015-1010.

[12] Ahn, C., Jeung, E.-B., 2023. Int. J. Mol. Sci. 24, 5342, https://doi.org/10.3390/ijms24065342.

[13] United Nations Environment Programme (UNEP), World Health Organisation (WHO), 2013. Bergman, A., Heindel, J.J., Jobling, S., Zoeller, R.T., editors. State of the science of endocrine disrupting chemicals – 2012. Geneva: United Nations Environment Programme, World Health Organization, https://www.who.int/publications/i/item/state-of-the-science-of-endocrine-disrupting-chemicals (last accessed on 06 October 2025).

[14] Toledano, J.M., Puche-Juarez, M., Moreno-Fernandez, J., Gonzalez-Palacios, P., Rivas, A., Ochoa, J.J., Diaz-Castro, J., 2024. Nutrients 16, 1556, https://doi.org/10.3390/nu16111556.

[15] Nesan, D., Kurrasch, D.M., 2020. Annu. Rev. Physiol. 82, 177–202, https://doi.org/10.1146/annurev-physiol-021119-034555.

[16] Braun, J.M., 2017. Nat. Rev. Endocrinol. 13, 161–173, https://doi.org/10.1038/nrendo.2016.186.

[17] Street, M.E., Bernasconi, S., 2020. Int. J. Mol. Sci. 21, 1430, https://doi.org/10.3390/ijms21041430.

[18] Dias, G.R.M., Giusti, F.C.V., de Novais, C.O., de Oliveira, M.A.L., Paiva, A.G., Kalil-Cutti, B., Mahoney, M.M., Graceli, J.B., 2025. Front. Endocrinol. 16:1571689, https://doi.org/10.3389/fendo.2025.1571689.

[19] Brehm, E., Flaws, J.A., 2019. Endocrinology 160, 1421–1435, https://doi.org/10.1210/en.2019-00034.

[20] Xin, F., Susiarjo, M., Bartolomei, M.S., 2015. Semin. Cell. Dev. Biol. 43, 66–75, https://doi.org/10.1016/j.semcdb.2015.05.008.

[21] Food and Agriculture Organisation of the United Nations (FAO), 2024. Exposure to endocrine disrupting chemicals – Changes from 2002 to 2024. Food Safety and Quality Series, No. 30, Rome, https://doi.org/10.4060/cd3005en.
[22] Priyadarshini, E., Parambil, A.M., Rajamani, P., Ponnusamy, V.K., Chen, Y.-H., 2023. Environ. Res. 225, 115577, https://doi.org/10.1016/j.envres.2023.115577.
[23] Rivera-Núñez, Z., Barrett, E.S., Szamreta, E.A., Shapses, S.A., Qin, B., Lin, Y., Zarbl, H., Buckley, B., Bandera, E.V., 2019. Environ. Health 18, 24, https://doi.org/10.1186/s12940-019-0464-8.
[24] Karthikeyan, B.S., Ravichandran, J., Mohanraj, K., Vivek-Ananth, R.P., Samal, A., 2019. Sci. Total Environ. 692, 281–296, https://doi.org/10.1016/j.scitotenv.2019.07.225.
[25] Metcalfe, C.D., Bayen, S., Desrosiers, M., Muñoz, G., Sauvé, S., Yargeau, V., 2022. Environ. Res. 207, 112658, https://doi.org/10.1016/j.envres.2021.112658.
[26] Peivasteh-roudsari, L., Barzegar-bafrouei, R., Sharifi, K.A., Azimisalim, S., Karami, M., Abedinzadeh, S., Asadinezhad, S., Tajdar-oranj, B., Mahdavi, V., Alizadeh, A.M., Sadighara, P., Ferrante, M., Conti, G.O., Aliyeva, A., Mousavi Khaneghah, A., 2023. Heliyon 9, e18140, https://doi.org/10.1016/j.heliyon.2023.e18140.
[27] Zhu, L., Hajeb, P., Fauser, P., Vorkamp, K., 2023. Sci. Total Environ. 874, 162374, https://doi.org/10.1016/j.scitotenv.2023.162374.
[28] Gonsioroski, A., Mourikes, V.E., Flaws, J.A., 2020. Int. J. Mol. Sci. 21, 1929 https://doi.org/10.3390/ijms21061929.
[29] Darbre, P.D., 2018. Int. J. Gen. Med. 11, 191–207, https://doi.org/10.2147/IJGM.S102230.
[30] Wong, K.H., Durrani, T.S., 2017. Curr. Probl. Pediatr. Adolesc. Health Care 47, 107–118, https://doi.org/10.1016/j.cppeds.2017.04.002.
[31] Prichystalova, R., Caron-Beaudoin, E., Richardson, L., Dirkx, E., Amadou, A., Zavodna, T., Cihak, R., Cogliano, V., Hynes, J., Pelland-St-Pierre, L., Verner, M.A., van Tongeren, M., Ho, V., 2021. J. Expo. Sci. Environ. Epidemiol. 31, 753–768, https://doi.org/10.1038/s41370-020-0253-z.
[32] Sellami, A., Reau, M., Montes, M., Lagarde, N., 2022. Front. Endocrinol. 13, 986016, https://doi.org/10.3389/fendo.2022.986016.
[33] La Merrill, M.A., Vandenberg, L.N., Smith, M.T., Goodson, W., Browne, P., Patisaul, H.B., Guyton, K.Z., Kortenkamp, A., Cogliano, V.J., Woodruff, T.J., Rieswijk, L., Sone, H., Korach, K.S., Gore, A.C., Zeise, L., Zoeller, R.T., 2020. Nat. Rev. Endocrinol. 16, 45–57, https://doi.org/10.1038/s41574-019-0273-8.
[34] Fekete, C., Lechan, R.M., 2014. Endocrine Reviews 35, 159–194, https://doi.org/10.1210/er.2013-1087.
[35] Bassett, J.H.D., Williams, G.R., 2016. Endocr. Rev. 37, 135–187, https://doi.org/10.1210/er.2015-1106.
[36] Ortiga-Carvalho, T.M., Chiamolera, M.I., Pazos-Moura, C.C., Wondisford, F.E., 2016. Compr. Physiol. 6, 1387–1428, https://doi.org/10.1002/cphy.c150027.
[37] Zoeller, R.T., Tan, S.W., Tyl, R.W., 2007. Crit. Rev. Toxicol. 37, 11–53, https://doi.org/10.1080/10408440601123446.
[38] Haigis, A.-C., Vergauwen, L., LaLone, C.A., Villeneuve, D.L., O'Brien, J.M., Knapen, D., 2023. Toxicol. Sci. 195, 1–27, https://doi.org/10.1093/toxsci/kfad063.
[39] Murk, A.J., Rijntjes, E., Blaauboer, B.J., Clewell, R., Crofton, K.M., Dingemans, M.M.L., David Furlow, J., Kavlock, R., Köhrle, J., Opitz, R., Traas, T., Visser, T.J., Xia, M., Gutleb, A.C., 2013. Toxicol. in Vitro 27, 1320–1346, https://doi.org/10.1016/j.tiv.2013.02.012.
[40] Brent, G.A., 2012. J. Clin. Invest. 122, 3035–3043, https://doi.org/10.1172/JCI60047.
[41] Schroeder, A.C., Privalsky, M.L., 2014. Front. Endocrinol. 5, 40, https://doi.org/10.3389/fendo.2014.00040.
[42] Mullur, R., Liu, Y.-Y., Brent, G.A., 2014. Physiol. Rev. 94, 355–382, https://doi.org/10.1152/physrev.00030.2013.
[43] Klein, I., Ojamaa, K., 2001. N. Eng. J. Med. 344, 501–509, https://doi.org/10.1056/NEJM200102153440707.

[44] Silva, J.F., Ocarino, N.M., Serakides, R., 2018. Biol. Reprod. 99, 907–921, https://doi.org/10.1093/biolre/ioy115.
[45] Williams, G.R., 2009. Endokrynol. Pol. 60, 380–388, https://doi.org/10.5603/ep.25453.
[46] Ravera, S., Nicola, J.P., Salazar-De Simone, G., Sigworth, F.J., Karakas, E., Amzel, L.M., Bianchet, M.A., Carrasco, N., 2022. Nature 612, 795–801, https://doi.org/10.1038/s41586-022-05530-2.
[47] Ruf, J., Carayon, P., 2006. Arch. Biochem. Biophys. 445, 269–277, https://doi.org/10.1016/j.abb.2005.06.023.
[48] Venugopalan, V., Al-Hashimi, A., Rehders, M., Golchert, J., Reinecke, V., Homuth, G., Völker, U., Manirajah, M., Touzani, A., Weber, J., Bogyo, M.S., Verrey, F., Wirth, E.K., Schweizer, U., Heuer, H., Kirstein, J., Brix, K., 2021. Int. Mol. J. Sci. 22, 462, https://doi.org/10.3390/ijms22010462.
[49] Richardson, S.J., Wijayagunaratne, R.C., D'Souza, D.G., Darras, V.M., Van Herck, S.L.J., 2015. Front. Neurosci. 9, https://doi.org/10.3389/fnins.2015.00066.
[50] Giammanco, M., Di Liegro, C.M., Schiera, G., Di Liegro, I., 2020. Int. Mol. J. Sci. 21, 4140, https://doi.org/10.3390/ijms21114140.
[51] Galton, V.A., 2017. Mol. Cell. Endocrinol. 458, 105–111, https://doi.org/10.1016/j.mce.2017.01.029.
[52] Gereben, B., Zavacki, A.M., Ribich, S., Kim, B.W., Huang, S.A., Simonides, W.S., Zeöld, A., Bianco, A.C., 2008. Endocr. Rev. 29, 898–938, https://doi.org/10.1210/er.2008-0019.
[53] Sandler, B., Webb, P., Apriletti, J.W., Huber, B.R., Togashi, M., Lima, S.T.C., Juric, S., Nilsson, S., Wagner, R., Fletterick, R.J., Baxter, J.D., 2004. J. Biol. Chem. 279, 55801–55808, https://doi.org/10.1074/jbc.M410124200.
[54] Mitra, T., Gulati, R., Ramachandran, K., Rajiv, R., Enninga, E.A.L., Pierret, C.K., Kumari R, S., Janardhanan, R., 2024. Diabetol. Metab. Syndr. 16, 95, https://doi.org/10.1186/s13098-024-01317-9.
[55] Alsen, M., Sinclair, C., Cooke, P., Ziadkhanpour, K., Genden, E., van Gerwen, M., 2021. Toxics 9, 1–26, https://doi.org/10.3390/toxics9010014.
[56] Gilbert, M.E., O'Shaughnessy, K.L., Axelstad, M., 2020. Endocrinology 161, bqaa106, https://doi.org/10.1210/endocr/bqaa106.
[57] Mughal, B.B., Fini, J.-B., Demeneix, B.A., 2018. Endocr. connect. 7, R160-R186, https://doi.org/10.1530/EC-18-0029.
[58] Crofton, K.M., 2008. Int. J. Androl. 31, 209–223, https://doi.org/10.1111/j.1365-2605.2007.00857.x.
[59] Kortenkamp, A., Axelstad, M., Baig, A.H., Bergman, Å., Bornehag, C.-G., Cenijn, P., Christiansen, S., Demeneix, B., Derakhshan, A., Fini, J.-B., Frädrich, C., Hamers, T., Hellwig, L., Köhrle, J., Korevaar, T.I.M., Lindberg, J., Martin, O., Meima, M.E., Mergenthaler, P., Nikolov, N., Du Pasquier, D., Peeters, R.P., Platzack, B., Ramhøj, L., Remaud, S., Renko, K., Scholze, M., Stachelscheid, H., Svingen, T., Wagenaars, F., Wedebye, E.B., Zoeller, R.T., 2020. Int. J. Mol. Sci. 21, 3123, https://doi.org/10.3390/ijms21093123.
[60] Noyes, P.D., Friedman, K.P., Browne, P., Haselman, J.T., Gilbert, M.E., Hornung, M.W., Barone, S., Crofton, K.M., Laws, S.C., Stoker, T.E., Simmons, S.O., Tietge, J.E., Degitz, S.J., 2019. Environ. Health Perspect. 127, 095001, https://doi.org/10.1289/EHP5297.
[61] Rodrigues, V.G., Henrique, G., Sousa-Vidal, É.K., de Souza, R.M.M., Tavares, E.F.C., Mezzalira, N., Marques, T. de O., Alves, B.M., Pinto, J.A.A., Irikura, L.N.N., Silva, R.E.C. da, de Oliveira, K.C., Maciel, R.M. de B., Giannocco, G., Serrano-Nascimento, C., 2024. Endocrines 5, 430–453, https://doi.org/10.3390/endocrines5030032.
[62] Sokal, A., Jarmakiewicz-Czaja, S., Tabarkiewicz, J., Filip, R., 2021. Nutrients 13, 867 https://doi.org/10.3390/nu13030867.
[63] Köhrle, J., Frädrich, C., 2021. Best Pract. Res. Clin. Endocrinol. Metab. 35, 101562, https://doi.org/10.1016/j.beem.2021.101562.
[64] Calsolaro, V., Pasqualetti, G., Niccolai, F., Caraccio, N., Monzani, F., 2017. Int. J. Mol. Sci. 18, 2583, https://doi.org/10.3390/ijms18122583.
[65] Boas, M., Feldt-Rasmussen, U., Main, K.M., 2012. Mol. Cell. Endocrinol. 355, 240–248, https://doi.org/10.1016/j.mce.2011.09.005.
[66] Boas, M., Feldt-Rasmussen, U., Skakkebæk, N.E., Main, K.M., 2006. Eur. J. Endocrinol. 154, 599–611, https://doi.org/10.1530/eje.1.02128.

[67] Brucker-Davis, F., 1998. Thyroid 8, 827–856, https://doi.org/10.1089/thy.1998.8.827.
[68] Puthiyachirakal, M.A., Hopkins, M., AlNatsheh, T., Das, A., 2025. Matern. Health Neonatol. Perinatol. 11, 9, https://doi.org/10.1186/s40748-025-00208-9.
[69] Segni, M., 2017. Disorders of the Thyroid Gland in Infancy, Childhood and Adolescence. South Dartmouth (MA): MDText.com, Inc.; 2000-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279032/ (last accessed on 06 October 2025).
[70] Alemu, A., Terefe, B., Abebe, M., Biadgo, B., 2016. Int. J. Reprod. Biomed. 14, 677-686, https://pmc.ncbi.nlm.nih.gov/articles/PMC5153572/.
[71] Moog, N.K., Entringer, S., Heim, C., Wadhwa, P.D., Kathmann, N., Buss, C., 2017. Neuroscience 342, 68–100, https://doi.org/10.1016/j.neuroscience.2015.09.070.
[72] Korevaar, T.I.M., Medici, M., Visser, T.J., Peeters, R.P., 2017. Nat. Rev. Endocrinol. 13, 610–622, https://doi.org/10.1038/nrendo.2017.93.
[73] Ghassabian, A., Trasande, L., 2018. Front. Endocrinol. 9:204, https://doi.org/10.3389/fendo.2018.00204.
[74] Delitala, A.P., Capobianco, G., Cherchi, P.L., Dessole, S., Delitala, G., 2019. Arch. Gynecol. Obstet. 299, 327–338, https://doi.org/10.1007/s00404-018-5018-8.
[75] Van Dingenen, I., Vergauwen, L., Haigis, A.-C., Blackwell, B.R., Stacy, E., Villeneuve, D.L., Knapen, D., 2023. Aquat. Toxicol. 261, 106632, https://doi.org/10.1016/j.aquatox.2023.106632.
[76] Baumann, L., Ros, A., Rehberger, K., Neuhauss, S.C.F., Segner, H., 2016. Aquat. Toxicol. 172, 44–55, https://doi.org/10.1016/j.aquatox.2015.12.015.
[77] Thambirajah, A.A., Koide, E.M., Imbery, J.J., Helbing, C.C., 2019. Front. Endocrinol. 10:276, https://doi.org/10.3389/fendo.2019.00276.
[78] Regnault, C., Usal, M., Veyrenc, S., Couturier, K., Batandier, C., Bulteau, A.-L., Lejon, D., Sapin, A., Combourieu, B., Chetiveaux, M., Le May, C., Lafond, T., Raveton, M., Reynaud, S., 2018. 115, E4416–E4425, https://doi.org/10.1073/pnas.1721267115.
[79] Pandey, S.P., Mohanty, B., 2017. NeuroToxicology 60, 16–22, https://doi.org/10.1016/j.neuro.2017.02.010.
[80] Sebastiano, M., Jouanneau, W., Blévin, P., Angelier, F., Parenteau, C., Pallud, M., Ribout, C., Gernigon, J., Lemesle, J.C., Robin, F., Pardon, P., Budzinski, H., Labadie, P., Chastel, O., 2023. Sci. Total Environ, 901, 165920, https://doi.org/10.1016/j.scitotenv.2023.165920.
[81] Yee, A.W., Aldeghi, M., Blakeley, M.P., Ostermann, A., Mas, P.J., Moulin, M., de Sanctis, D., Bowler, M.W., Mueller-Dieckmann, C., Mitchell, E.P., Haertlein, M., de Groot, B.L., Boeri Erba, E., Forsyth, V.T., 2019. Nat. Commun. 10, 925, https://doi.org/10.1038/s41467-019-08609-z.
[82] Sanguinetti, C., Minniti, M., Susini, V., Caponi, L., Panichella, G., Castiglione, V., Aimo, A., Emdin, M., Vergaro, G., Franzini, M., 2022. Biomedicines 10, 1906, https://doi.org/10.3390/biomedicines10081906.
[83] Richardson, S.J., 2009. FEBS J. 276, 5342–5356, https://doi.org/10.1111/j.1742-4658.2009.07244.x.
[84] Saha, S., Chakraborty, S., Bhattacharya, A., Biswas, A., Ain, R., 2017. Sci. Rep. 7, 16548, https://doi.org/10.1038/s41598-017-16566-0.
[85] Alshehri, B., D'Souza, D.G., Lee, J.Y., Petratos, S., Richardson, S.J., 2015. J. Neuroendocrinol. 27, 303–323, https://doi.org/10.1111/jne.12271.
[86] Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Environ. Toxicol. Chem. 29, 730–741, https://doi.org/10.1002/etc.34.
[87] Li, S., Qin, S., Zeng, H., Chou, W., Oudin, A., Kanninen, K.M., Jalava, P., Dong, G., Zeng, X., 2024. Eco-Environ. Health 3, 476-493, https://doi.org/10.1016/j.eehl.2024.08.002.
[88] Dierichs, N.T.O.M., Piersma, A.H., Peeters, R.P., Visser, W.E., Meima, M.E., Hessel, E.V.S., 2025. Crit. rev. Toxicol. 55, 304–320, https://doi.org/10.1080/10408444.2025.2461076.
[89] Gadaleta, D., Spînu, N., Roncaglioni, A., Cronin, M.T.D., Benfenati, E., 2022. Int. J. Mol. Sci. 23, 3053, https://doi.org/10.3390/ijms23063053.
[90] Spinu, N., Bal-Price, A., Cronin, M.T.D., Enoch, S.J., Madden, J.C., Worth, A.P., 2019. Arch. Toxicol. 93, 2759–2772. https://doi.org/10.1007/s00204-019-02551-1.

[91] European Commission, 1999. Community Strategy for Endocrine Disrupters - a range of substances suspected of interfering with the hormone systems of humans and wildlife. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1999:0706:FIN:EN:PDF (last accessed on 06 October 2025).
[92] European Commission, 2019. The European Green Deal. https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (last accessed on 06 October 2025).
[93] European Commission, 2020. Chemicals Strategy for Sustainability towards a Toxic-free Environment. https://environment.ec.europa.eu/strategy/chemicals-strategy_en (last accessed on 06 October 2025).
[94] REACH Regulation (EC) No. 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals. Brussels, Belgium., 2006.
[95] European Commission, 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.
[96] European Commission, 2012. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products (Text with EEA relevance).
[97] European Commission, Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (Text with EEA relevance).
[98] Burden, N., Brown, R.J., Smith, R., Brescia, S., Goodband, T., Guerrero-Limón, G., Kent, L., Marty, S., Pearson, A., van der Mescht, M., Saunders, L.J., Sewell, F., Wang, N., Wheeler, J.R., 2024. Regul. Toxicol. Pharmacol. 151, 105671. https://doi.org/10.1016/j.yrtph.2024.105671.
[99] European Commission, 2018. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties (Text with EEA relevance).
[100] European Commission, 2017. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council (Text with EEA relevance).
[101] European Commission, 2022. Commission Delegated Regulation (EU) 2023/707 of 19 December 2022 amending Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures (Text with EEA relevance).
[102] Holmer, M.L., Holmberg, R.D., Despicht, C., Bouftas, N., Axelstad, M., Beronius, A., Zilliacus, J., Van Duursen, M., Svingen, T., 2025. Regul. Toxicol. Pharmacol. 162, 105883, https://doi.org/10.1016/j.yrtph.2025.105883.
[103] Andersson, N., Arena, M., Auteri, D., Barmaz, S., Grignard, E., Kienzler, A., Lepper, P., Lostia, A.M., Munn, S., Parra Morte, J.M., Pellizzato, F., Tarazona, J., Terron, A., Van der Linden, S., 2018. EFSA J. 16, e05311, https://doi.org/10.2903/j.efsa.2018.5311.
[104] European Chemicals Agency (ECHA), 2025. Key Areas of Regulatory Challenge. https://doi.org/10.2823/8572710.
[105] Burden, N., Mitchell, C.A., Dang, Z., Embry, M.R., Glaberman, S., Lagadic, L., Lynn, S.G., Marini, J., Mihaich, E., McDermott, E., Krzykwa, J., Salinas, E.R., Schoenfuss, H., Thorpe, K., Weltje, L., Wheeler, J.R., 2025. Environ. Toxicol. Chem. 44, 1477–1496, https://doi.org/10.1093/etojnl/vgaf064.
[106] Worth, A.P., Berggren, E., Prieto, P., 2025. Chemicals 2.0 and Why We Need to Bypass the Gold Standard in Regulatory Toxicology. Altern. Lab. Anim. 53, 21–25, https://doi.org/10.1177/02611929241296328.
[107] Schmeisser, S., Miccoli, A., von Bergen, M., Berggren, E., Braeuning, A., Busch, W., Desaintes, C., Gourmelon, A., Grafström, R., Harrill, J., Hartung, T., Herzler, M., Kass, G.E.N., Kleinstreuer, N., Leist, M., Luijten, M., Marx-Stoelting, P., Poetz, O., van Ravenzwaay, B., Roggeband, R., Rogiers,

V., Roth, A., Sanders, P., Thomas, R.S., Marie Vinggaard, A., Vinken, M., van de Water, B., Luch, A., Tralau, T., 2023. Environ. Int. 178, 108082, https://doi.org/10.1016/j.envint.2023.108082.

[108] Fischer, I., Milton, C., Wallace, H., 2020. Toxicol. Res. 9, 67–80, https://doi.org/10.1093/toxres/tfaa011.

[109] Walder, L., Pallocca, G., Bastos, L.F., Beekhuijzen, M., Busquet, F., Constantino, H., Corvaro, M., Courtot, L., Escher, B., Fernandez, R., Gougeon, E., Hansell, L., Herzler, M., Holden, L., Hornek-Gausterer, R., Irizar, A., Kandarova, H., Kern, P., Kolle, S., Lacasse, K., Lee, I., Macmillan, D.S., Maxwell, G., Moriarty, O., Nadzialek, S., Pochat, J., Reid, K., Revel, M., Ritskes-Hoitinga, M., Sobanski, T., Stoddart, G., Underhill, D., Veillette, M., Vriend, J., Westmoreland, C., Baines, J., 2025. ALTEX 42, 435–450 https://doi.org/10.14573/altex.2503241.

[110] Stoykova, K., 2025. Eur. J. Risk Regul. 1–32, https://doi.org/10.1017/err.2025.1.

[111] Herzler, M., Luijten, M., Marx-Stoelting, P., Rivière, G., 2025. Curr. Opin. Toxicol. 42, 100517, https://doi.org/10.1016/j.cotox.2025.100517.

[112] Berggren, E., Worth, A.P., 2023. Regul. Toxicol. Pharmacol. 142, 105431, https://doi.org/10.1016/j.yrtph.2023.105431.

[113] Bearth, A., Roth, N., Jansen, T., Holden, L., Čavoški, A., Di Consiglio, E., Hauzenberger, I., Lee, R., Mombelli, E., Tcheremenskaia, O., Wendt-Rasch, L., Wilks, M.F., 2025. Environ. Int. 196, 109279, https://doi.org/10.1016/j.envint.2025.109279.

[114] Browne, P., Van Der Wal, L., Gourmelon, A., 2020. Mol. Cell. Endocrinol. 504, 110675, https://doi.org/10.1016/j.mce.2019.110675.

[115] Svingen, T., Schwartz, C.L., Rosenmai, A.K., Ramhøj, L., Johansson, H.K.L., Hass, U., Draskau, M.K., Davidsen, N., Christiansen, S., Ballegaard, A.-S.R., Axelstad, M., 2022. Environ. Poll. 304, 119242, https://doi.org/10.1016/j.envpol.2022.119242.

[116] Wiklund, L., Pípal, M., Weiss, J., Beronius, A., 2024. Toxicology 504, 153794, https://doi.org/10.1016/j.tox.2024.153794.

[117] van der Zalm, A.J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T.R., Kleinstreuer, N.C., Lowit, A.B., Perron, M., Clippinger, A.J., 2022. Arch. Toxicol. 96, 2865–2879 https://doi.org/10.1007/s00204-022-03365-4.

[118] Bajard, L., Adamovsky, O., Audouze, K., Baken, K., Barouki, R., Beltman, J.B., Beronius, A., Bonefeld-Jørgensen, E.C., Cano-Sancho, G., de Baat, M.L., Di Tillio, F., Fernández, M.F., FitzGerald, R.E., Gundacker, C., Hernández, A.F., Hilscherova, K., Karakitsios, S., Kuchovska, E., Long, M., Luijten, M., Majid, S., Marx-Stoelting, P., Mustieles, V., Negi, C.K., Sarigiannis, D., Scholz, S., Sovadinova, I., Stierum, R., Tanabe, S., Tollefsen, K.E., van den Brand, A.D., Vogs, C., Wielsøe, M., Wittwehr, C., Blaha, L., , 2023. Environ. Res. 217, 114650, https://doi.org/10.1016/j.envres.2022.114650.

[119] Organisation for Economic Co-operation and Development (OECD), 2018. Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. https://doi.org/10.1787/9789264304741-1-en.

[120] Mitchell, C.A., Burden, N., Bonnell, M., Hecker, M., Hutchinson, T.H., Jagla, M., LaLone, C.A., Lagadic, L., Lynn, S.G., Shore, B., Song, Y., Vliet, S.M., Wheeler, J.R., Embry, M.R., 2023. Environ. Toxicol. Chem. 42, 757–777, https://doi.org/10.1002/etc.5584.

[121] Ramhøj, L., Axelstad, M., Baert, Y., Cañas-Portilla, A.I., Chalmel, F., Dahmen, L., De La Vieja, A., Evrard, B., Haigis, A.-C., Hamers, T., Heikamp, K., Holbech, H., Iglesias-Hernandez, P., Knapen, D., Marchandise, L., Morthorst, J.E., Nikolov, N.G., Nissen, A.C.V.E., Oelgeschlaeger, M., Renko, K., Rogiers, V., Schüürmann, G., Stinckens, E., Stub, M.H., Torres-Ruiz, M., Van Duursen, M., Vanhaecke, T., Vergauwen, L., Wedebye, E.B., Svingen, T., 2023. Front. Toxicol. 5, 1189303, https://doi.org/10.3389/ftox.2023.1189303.

[122] Barton-Maclaren, T.S., Wade, M., Basu, N., Bayen, S., Grundy, J., Marlatt, V., Moore, R., Parent, L., Parrott, J., Grigorova, P., Pinsonnault-Cooper, J., Langlois, V.S., 2022. Environ. Res. 204, 112225, https://doi.org/10.1016/j.envres.2021.112225.

[123] Muratov, E.N., Bajorath, J., Sheridan, R.P., Tetko, I.V., Filimonov, D., Poroikov, V., Oprea, T.I., Baskin, I.I., Varnek, A., Roitberg, A., Isayev, O., Curtalolo, S., Fourches, D., Cohen, Y., Aspuru-

Guzik, A., Winkler, D.A., Agrafiotis, D., Cherkasov, A., Tropsha, A., 2020. Chem. Soc. Rev. 49, 3525–3564, https://doi.org/10.1039/D0CS00098A.
[124] Benigni, R., 2017. Appl. In Vitro Toxicol. 3, 265–270, https://doi.org/10.1089/aivt.2017.0007.
[125] Leist, M., Ghallab, A., Graepel, R., Marchan, R., Hassan, R., Bennekou, S.H., Limonciel, A., Vinken, M., Schildknecht, S., Waldmann, T., Danen, E., van Ravenzwaay, B., Kamp, H., Gardner, I., Godoy, P., Bois, F.Y., Braeuning, A., Reif, R., Oesch, F., Drasdo, D., Höhme, S., Schwarz, M., Hartung, T., Braunbeck, T., Beltman, J., Vrieling, H., Sanz, F., Forsby, A., Gadaleta, D., Fisher, C., Kelm, J., Fluri, D., Ecker, G., Zdrazil, B., Terron, A., Jennings, P., van der Burg, B., Dooley, S., Meijer, A.H., Willighagen, E., Martens, M., Evelo, C., Mombelli, E., Taboureau, O., Mantovani, A., Hardy, B., Koch, B., Escher, S., van Thriel, C., Cadenas, C., Kroese, D., van de Water, B., Hengstler, J.G., 2017. Arch. Toxicol. 91, 3477–3505, https://doi.org/10.1007/s00204-017-2045-3.
[126] Allen, T.E.H., Goodman, J.M., Gutsell, S., Russell, P.J., 2016. Chem. Res. Toxicol. 29, 2060–2070, https://doi.org/10.1021/acs.chemrestox.6b00341.
[127] Organisation for Economic Co-operation and Development (OECD), 2021. Guidance Document for the scientific review of Adverse Outcome Pathways. https://doi.org/10.1787/a6bec14b-en.
[128] Holbech, H., Matthiessen, P., Hansen, M., Schüürmann, G., Knapen, D., Reuver, M., Flamant, F., Sachs, L., Kloas, W., Hilscherova, K., Leonard, M., Arning, J., Strauss, V., Iguchi, T., Baumann, L., 2020. Int. J. Mol. Sci. 21, 2954, https://doi.org/10.3390/ijms21082954.
[129] Audouze, K., Sarigiannis, D., Alonso-Magdalena, P., Brochot, C., Casas, M., Vrijheid, M., Babin, P.J., Karakitsios, S., Coumoul, X., Barouki, R., 2020. Int. J. Mol. Sci. 21, 2988, https://doi.org/10.3390/ijms21082988.
[130] Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics. Wiley, https://doi.org/10.1002/9783527628766.ch20.
[131] Pirhadi, S., Shiri, F., Ghasemi, J.B., 2015. RSC Adv. 5, 104635–104665, https://doi.org/10.1039/C5RA10729F.
[132] Organisation for Economic Co-operation and Development (OECD), 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en.
[133] Organisation for Economic Co-operation and Development (OECD), 2023. (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions. https://doi.org/10.1787/d96118f6-en.
[134] Gissi, A., Tcheremenskaia, O., Bossa, C., Battistelli, C.L., Browne, P., 2024. Comput. Toxicol. 31, 100326, https://doi.org/10.1016/j.comtox.2024.100326.
[135] Walker, J.D., Jaworska, J., Comber, M.H.I., Schultz, T.W., Dearden, J.C., 2003. Environ. Toxicol. Chem. 22, 1653–1665, https://doi.org/10.1897/01-627.
[136] Tropsha, A., 2010. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inform. 29, 476–488, https://doi.org/10.1002/minf.201000061.
[137] Weininger, D., 1988. J. Chem. Inf. Comput. Sci. 28, 31–36, https://doi.org/10.1021/ci00057a005.
[138] Wold, S., Esbensen, K., Geladi, P., 1987. Chemometr. Intell. Lab. 2, 37–52, https://doi.org/10.1016/0169-7439(87)80084-9.
[139] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second edition, Springer: New York, NY, USA, 2009. https://doi.org/10.1007/978-0-387-84858-7.
[140] Leardi, R., Boggia, R., Terrile, M., 1992. J. Chemometr. 6, 267–281, https://doi.org/10.1002/cem.1180060506.
[141] Rücker, C., Rücker, G., Meringer, M., 2007. J. Chem. Inf. Model. 47, 2345–2357, https://doi.org/10.1021/ci700157b.
[142] Todeschini, R., Consonni, V., Maiocchi, A., 1999. Chemometr. Intell. Lab. 46, 13–29, https://doi.org/10.1016/S0169-7439(98)00124-5.
[143] Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. J. Comput. Chem. 34, 2121–2132, https://doi.org/10.1002/jcc.23361.

[144] 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., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Altern. Lab. Anim. 33, 155–173, https://doi.org/10.1177/026119290503300209.
[145] Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., Todeschini, R., 2012. Molecules 17, 4791–4810, https://doi.org/10.3390/molecules17054791.
[146] Klingspohn, W., Mathea, M., ter Laak, A., Heinrich, N., Baumann, K., 2017. J. Cheminform. 9, 44, https://doi.org/10.1186/s13321-017-0230-2.
[147] Mathea, M., Klingspohn, W., Baumann, K., 2016. Mol. Inform. 35, 160–180, https://doi.org/10.1002/minf.201501019.
[148] Faraway, Julian J, 2002. Practical Regression and Anova using R. https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf (last accessed on 06 October 2025).
[149] ACD ChemSketch software, ACDLabs, Toronto, ON, Canada, https://www.acdlabs.com/ (last accessed on 06 October 2025).
[150] O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. J. Cheminform. 3, 33, https://doi.org/10.1186/1758-2946-3-33.
[151] Yap, C.W., 2011. J. Comput. Chem. 32, 1466–1474, https://doi.org/10.1002/jcc.21707.

CHAPTER 2

In Silico Models for the Screening of Human Transthyretin Disruptors

Marco Evangelista, Nicola Chirico, and Ester Papa

Published in Journal of Hazardous Materials, Volume 480, 136188, 2024, https://doi.org/10.1016/j.jhazmat.2024.136188

Abstract

The use of new approach methodologies (NAMs), such as quantitative structure-activity relationship (QSAR) models, is highly recommended by international regulations to speed up hazard and risk assessment of endocrine disruptors, which are known to be linked to a wide spectrum of severe diseases on humans and wildlife. A very sensitive target for these chemicals is the thyroid hormone system, which plays a key role in regulating metabolic and cognitive functions. Several chemicals have been demonstrated to compete with the thyroid hormone thyroxine (T4) for binding to human thyroid hormone distributor protein transthyretin (hTTR). In this work, we generated three new datasets composed by T4-hTTR competing potencies of more than 200 heterogeneous chemicals measured by three different *in vitro* assays. These datasets were used for the development of new regression QSAR models. The best models were thoroughly validated by internal and external validation procedures. The mechanistic interpretation of the selected molecular descriptors provided information on structural features which are relevant to characterise hTTR binders, such as the presence of hydroxylated and halogenated aromatic rings. Principal component analysis was used to rank the studied chemicals according to their increasing T4-hTTR competing potency. Hydroxylated and halogenated bicyclic aromatic compounds are ranked as the strongest hTTR binders. The new QSARs are useful to screen potential thyroid hormone system-disrupting chemicals, and to support the identification of sustainable alternatives to hazardous chemicals.

2.1 Introduction

The concern about endocrine disrupting chemicals (EDCs) is significantly increasing worldwide, owing to spreading evidences of endocrine-related disorders both in human and wildlife, such as reduction of fertility, cancers, metabolic and developmental dysfunctions [1,2]. According to the World Health Organization International Programme on Chemical Safety (WHO/IPCS), an endocrine disruptor is defined as "an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations" [3]. The risk assessment of EDCs is extremely challenging [4,5] due to their structural heterogeneity and large variety of physicochemical properties and environmental behaviours, their multiple emission sources, and the complexity of the endocrine system [3]. These features, on the one hand, lead to the ubiquitous presence of EDCs in the environmental compartments, increasing the number and complexity of the potential exposure pathways, while, on the other hand, determine several modes of action and toxicity pathways by which those chemicals exert their adverse effects on living organisms [6]. In Europe, several political and scientific efforts followed the Community Strategy for Endocrine Disruptors [7], aiming to fill knowledge gaps, to set validated test methods and to harmonize the current chemical legislations [8]. A priority of the recent European Green Deal [9] is to improve the protection of human health and of the environment against harmful chemicals, moving towards a toxic-free environment through a Chemical Strategy for Sustainability (CSS) [10], by strengthening and harmonizing the European regulatory system, with a significant focus on endocrine disruptors such as in the recent Classification, Labelling and Packaging (CLP) Regulation update [11].

Among EDCs, a variety of industrial chemicals, pesticides, plasticizers, surfactants, pharmaceuticals and personal care products [12,15] have been recognised to interfere with multiple molecular targets involved in the hypothalamic-pituitary-thyroid (HPT) axis through several mechanisms, altering the thyroid hormones (THs) homeostasis [16,17]. Alterations of the TH homeostasis may induce severe adverse diseases to nervous system development, metabolism, immune and cardiovascular systems [18,19]. Substances able to exert this type of activity are known as thyroid hormone system-disrupting chemicals (THSDCs) [20,21]. Triiodothyronine (T3) and thyroxine (T4) represent the two main hormones produced by the thyroid gland, from where THs are transported into the bloodstream by the activity of THs transporters (e.g., monocarboxylate transporter 8 or MCT8) [20,22]. In mammals, T4 is the predominant

TH form secreted by the thyroid, while T3 is the more biologically active one, due to its higher binding affinity with THs nuclear receptors (TRs). T4 is often referred to as a "prohormone", as it is converted to T3 in peripheral tissues through deiodination catalysed by deiodinase enzymes [23,24]. Nearly all the amount of THs in human blood is bound to distributor proteins, such as transthyretin (TTR), thyroxine-binding globulin (TBG), and albumin [25]. Differently from TBG and albumin, which are primarily synthetised by the liver, TTR is also synthetised by the choroid plexus, by the retinal pigment epithelium of the eye, and by other tissues and/or organs, implying that TTR is involved in multiple biological processes [26,27]. Among these different functions, TTR has the role to buffer abnormal alterations of free THs physiological concentrations. Free THs are devoted to reach the target tissues to guarantee the proper functioning of the TH system [28]. Any alteration in the THs binding to distributor proteins can modulate the free THs physiological concentrations in blood, leading to TH system dysfunctions. Particular attention is posed in studying the ability of THSDCs to displace the thyroid hormone T4 from the TTR. Differently from the other TH distributor proteins, TTR has a key role in delivering T4 in cerebrospinal fluid [18] across the blood-brain barrier and placenta during foetal development: exogenous chemicals able to bind to TTR could be transferred to the foetus, leading to a T4 deficiency and subsequent severe, irreversible effects, primarily cognitive dysfunctions [19,29]. For these reasons, the identification of chemicals with this behaviour is considered a priority [20,22].

In the literature, several chemical classes of compounds have been experimentally demonstrated to compete with T4 for binding to the human TTR (hTTR), such as per- and polyfluoroalkyl substances (PFAS) [30], bisphenols [31], halogenated phenols [32], polychlorinated biphenyls (PCBs) [33], polybrominated diphenyl ethers (PBDEs) [34] and both PCBs and PBDEs hydroxylated metabolites [35,36]. The traditional approach based on *in vivo* experiments became inadequate and unsustainable for testing the potential endocrine properties of all the existing substances, and so for the identification of potential THSDCs [20,37]. In particular, the development of new approach methodologies (NAMs), such as *in vitro* test (e.g. *in vitro* assays to detect estrogen [38] and androgen [39] receptors agonists and antagonists), and *in silico* methods (e.g. quantitative structure-activity relationships (QSARs)), is highly recommended in the literature, and by several authorities, to support the identification of EDCs, as well as of THSDCs [8,20,37,40–42]. Worldwide, several efforts and projects aimed for an integration of the results in the weight of evidence (WoE) approach [43–46]. Focusing on the prediction of the hTTR binding potencies of chemicals, a limited

number of regression QSARs have been developed so far using techniques such as multiple linear regression (MLR) [32,47–50], k-nearest neighbors (k-NN) [50], and 3D QSAR approaches [51,52]. Most of these QSARs are based on small datasets (between 15 and 32 chemicals) mainly representative for specific chemical classes (such as PBDEs, perfluorinated compounds (PFCs), halogenated phenols and thiophenols). Among the aforementioned models, only those developed by Yang and coworkers [50] have a wide applicability domain in terms of structural heterogeneity represented in their training sets. Statistical performances of the existing regression models range, in terms of fitting (R^2), between 0.810 and 0.960. However, with the exclusion of the models by Yang and coworkers [50], the structural and response domains of these QSARs is very narrow due to the limited size of the respective training sets. Furthermore, the use of commercial software for the calculation of the molecular descriptors, such as the Dragon software, may limit their application.

This study aims to develop new regression QSAR models, following the OECD guidelines for QSAR development [53], for the identification of potential hTTR binders. This work takes into account current research shortcomings [41,54] and in particular the need for new QSAR models for endocrine disruption modalities different than androgen and estrogen (e.g. here thyroid hormone system disruption), based on curated datasets and with broad applicability domains. To this end, the QSARs proposed in this work were developed using three new and curated datasets based on data collected from the literature, which include values of T4-hTTR competing potencies for more than 200 different chemicals, expressed as relative competitive potency (RP), and calculated from experimental binding affinities measured using three *in vitro* assays (the 8-anilino-1-naphthalenesulfonic acid (ANSA) based binding assay [20], the fluorescence conjugate isothiocyanate (FITC)-T4 based binding assay [20], and the radiolabeled [125I]-T4 binding assay (RLBA) [34]). The development of models, based on these datasets, extends the experimental and chemical information contained in previous QSAR studies for the same endpoint. Furthermore, since the three datasets cover a wide range of different chemicals, each QSAR is applicable to fill the experimental data gaps for ANSA, FITC-T4, and RLBA assays. The experimental and predicted RP values for more than 200 chemicals were used to screen potential hTTR binders using the QSAR models developed for the different *in vitro* assays, to evaluate their agreement. A final important aspect that was considered in this study was the need to ease the reproducibility and the application of the new QSARs by implementing them in a non-commercial software.

2.2 Material and methods

2.2.1 Datasets

The raw dataset reported in the Supplementary Material S1 lists binding affinity (commonly quantified in terms of IC50, i.e. half-maximal inhibitory concentration) or RP values measured for 240 different chemicals, mostly aromatic and halogenated. This collection includes, to the best of our knowledge, all the binding affinity or RP values currently available in the literature [29,30,32–36,55–86]. RP is defined as the ratio between the binding affinity of T4 and the binding affinity of a chemical with hTTR. When only binding affinity values were available in the original literature, their RP values were calculated and included in the raw dataset (Supplementary Material S1). The use of RP values to define the hTTR binding affinity reduces the data bias caused by minor modifications of the procedures within the same *in vitro* assays, and by their measurements in different laboratories [84]. Data for mixtures, ambiguous molecular structures, and unclear chemical identifiers, were not included in the raw dataset. Moreover, in order to avoid further reduction of the already limited structural and experimental information, two chiral forms of the same chemical (hexabromocyclodecane alpha and beta) were converted in the non-chiral form, while structures of salts and anions (11 chemicals) were converted to their respective acid, as in a previous work [87]. Geometric mean of RP values was calculated when multiple experimental data were available for one chemical [88].

Three datasets (i.e. ANSA, FITC-T4, and RLBA), reported in Supplementary Material S2, were extracted from the raw dataset to develop QSAR models for the prediction of T4-hTTR competing potencies, according to different *in vitro* assays explained as follows. The RLBA, firstly described by Somack et al. [89], and further elaborated by minor modifications [34,35], uses radioactive iodine-125 (125I) to label T4 in order to make T4 detectable. A gamma counter is used to measure the radioactivity of the radiolabeled-T4 bound with TTR. Variation in the measured signal in the presence of a competitor is used to quantify the relative binding [34,35,89]. The ANSA-based binding assay (firstly described by Nilsson et al. [90], and further elaborated by minor modifications [55,71]), and the FITC-T4-based binding assay (firstly described by Smith et al. [91], and further elaborated by minor modifications [36,72]) are two competitive fluorescence displacement assays: based on different principles, the TTR binding of chemicals is quantified by measuring the degree of reduction of the

fluorescence signal [20]. Particularly, competitive fluorescence displacement assays have been highlighted as powerful methods to detect chemicals able to interfere with TH distributor proteins TTR and TBG [84], and have been recently used in the Joint Research Centre (JRC)'s EURL ECVAM thyroid validation study for the validation of methods for THSDCs identification [20].

The raw dataset also includes 35 binding affinities measured with surface plasmon resonance-based bioassay (SPRB) [69] and isothermal titration calorimetry (ITC) [85], reported in Supplementary Materials S1. However, these data were too limited for modelling purposes and were excluded from further analysis.

Dataset ANSA includes 79 compounds, such as hydroxylated PBDEs, hydroxylated PCBs, sulfated PCBs, phosphates, fatty acids, halogenated phenols and thiophenols, and halogenated and/or hydroxylated benzoic acids. Dataset FITC-T4 is composed by 50 compounds, such as PFAS, bisphenols, hydroxylated PBDEs, PCBs, hydroxylated PCBs, sulfated PCBs, parabens, phthalates, and halogenated phenols. Dataset RLBA includes 137 compounds, such as PBDEs, hydroxylated PBDEs, PCBs, hydroxylated PCBs, PFAS, halogenated phenols, bisphenol A derivatives, and xanthones. RP values were log transformed prior to QSAR modelling. Finally, all the experimental LogRP values from the ANSA, FITC-T4, and RLBA datasets, were combined into Dataset I, which includes 223 chemicals. Missing experimental LogRPs were filled by applying the respective assay-specific QSAR model developed in this work. This procedure led to the development of Dataset II, which includes 3 assay-specific columns of experimental and, where not available, predicted LogRP values. Finally, only predicted LogRPs within the range of the experimental response of the respective training set, with a prediction uncertainty below the maximum uncertainty predicted for the respective training set, and calculated for chemicals within the structural applicability domain of each model, were included into Dataset III (150 unique chemicals). Principal component analysis (PCA) [92] was applied to evaluate the relationships between the LogRP values within Dataset III. Dataset I, II, and III, are reported in Supplementary Material S2.

2.2.2 Molecular descriptors

The chemical structures were coded as SMILES (simplified molecular input line entry system), which were downloaded from the PubChem website (available at https://pubchem.ncbi.nlm.nih.gov/) or calculated by drawing the chemical structures

in the ACD/ChemSketch software v. 2021.1.1 (available at https://www.acdlabs.com/, ACDLabs, Toronto, ON, Canada). SMILES were harmonized using the Open Babel software v. 2.4.1 (available at https://openbabel.org/) [93] to generate unique strings for the same molecular structure, and were used as input by the PaDEL-Descriptor software v. 2.21 [94] for the calculation of 7185 mono- /bi- dimensional theoretical molecular descriptors and fingerprints, which encode for the structural features of the compounds. Molecular descriptors were filtered using the QSARINS software [95] prior to modelling, in order to reduce useless and redundant information. Descriptors with at least one missing value, as well as those characterized by low variance (i.e., the same value for more than 80 % of the molecules) or a pairwise correlation exceeding 95 %, were excluded from the further analysis. At the end of this procedure, 546, 483 and 503 molecular descriptors were selected for the datasets i.e., ANSA, FITC-T4 and RLBA, respectively.

2.2.3 Splitting

To estimate the ability of the QSAR models to make reliable predictions, each dataset was split in a training set for the model development, and in a test set for the external validation. Chemicals were sorted by response, then one every three chemicals were assigned to the test set, and the remaining chemicals were assigned to the training set. The first and the last chemicals were assigned to the training set, in order to limit the endpoint experimental space of the model and to reduce extrapolation. Finally, each training set was further pre-filtered using QSARINS by removing low variance and highly correlated descriptors, as described in Section 2.2.2.

2.2.4 Modelling procedure

QSARs were developed using MLR by means of ordinary least squares (OLS), and several statistical metrics were calculated to verify their goodness-of-fit (coefficient of determination, R^2, and root mean square error of the training set, $RMSE_{TR}$), internal robustness (leave-one-out, Q^2_{loo}, and leave-more-out, Q^2_{lmo}, which are cross validation metrics), and external predictive ability ($RMSE_{TEST}$, Q^2_{F3}). Details about the calculation of these metrics are reported in Supplementary Material S3.

The descriptors selection procedure applied to the filtered descriptors (see Section 2.2.3) explored, in a first step, all their combinations (all-subset) up to two descriptors; in a second step, a genetic algorithm (GA) was applied to explore the most promising

combinations of three or more descriptors [95]. Q^2_{LOO} was used as the fitness function both for all-subset and GA.

Moreover, to further check for the QSAR models robustness, both the QUIK rule [96] and the Y-scrambling procedures [97] were applied. The QUIK rule was performed to ensure that the correlations among the modelling descriptors is lower than their correlation with the response (in this work, models were filtered according to $K_{XY} - K_{XX} > 0.015$ [96]). To check for chance correlation between the molecular descriptors and the response of the developed QSARs, 2000 QSARs were developed using training sets with random shuffled responses (Y-scrambling). The average of R^2 and Q^2_{LOO} (called R^2_{YS} and Q^2_{YS}, respectively) of the 2000 QSARs should be much lower than the ones of the QSARs under scrutiny, since the structure-activity relationship of the scrambled QSARs should be negligible.

Calculations concerning the QSARs and the corresponding performances were performed using the QSARINS software [95].

2.2.5 Applicability domain

The definition of the applicability domain (AD) of a QSAR model is required to evaluate the reliability of predictions and/or the degree of extrapolation. Each QSAR model was developed using the structural and the experimental information included in the respective training set, which defines the space of the AD. Concerning the chemical structures, the AD of each model was defined by the leverage approach, with the graphical support of the Williams plot. In particular, this charts plots the standardized residuals of predictions (a measure of the response AD) on the y-axis, allowing for the identification of response outliers (i.e., chemicals with standardized residuals that fall outside the range defined by ±2.5 standard deviation units), and the leverage values (HAT matrix diagonal elements) of the chemicals on the x-axis, allowing for the identification of structural outliers (i.e., chemicals with a leverage value greater than h^*, defined as $3 *(p + 1)/n$, where p is the number of model descriptors and n is the number of the chemicals in the training set). The HAT matrix, also known as leverage or influence matrix, is calculated as

$$HAT = X(X^T X)^{-1} X^T \quad (2.1)$$

where X is the data matrix consisting of n rows and p columns.

The leverage value of each chemical is a measure of its distance from the centroid of the model. It quantifies the influence of each chemical on the model and the reliability of the predictions. Predictions associated to hat values larger than h^* (defined above) are considered outside the structural AD of the model and, therefore, are less reliable than those falling within the h^* cut-off value. The inclusion of predictions for new chemicals within the AD of the here proposed models can be verified in the QSAR-ME Profiler beta version 1.02 software available at https://dunant.dista.uninsubria.it/qsar/.

To further evaluate the QSAR models, the normality of the distribution of the residuals was graphically inspected through the QQ plots (quantile–quantile plots).

2.2.6 Prediction of missing LogRPs in Dataset I

The prediction of missing experimental LogRP values of chemicals included in Dataset I was performed by applying the QSAR models reported in equations 2.3, 2.5 and 2.7, as described in the Section 2.2.1, using the QSAR-ME Profiler beta version 1.02 software (available at https://dunant.dista.uninsubria.it/qsar/) leading to Dataset II. This software allowed for the evaluation of the reliability of predictions by identifying the chemicals within the applicability domains, and by comparing the uncertainty of predictions with the range of the uncertainty in the training sets. This procedure led to the development of Dataset III, as described in Section 2.2.1. Predicted LogRPs, hat values, and uncertainties values, are reported in Supplementary Material S2. Additional details on the calculation of uncertainties in the training and in the test set are reported in Supplementary Material S3.

2.3 Results and discussion

The ANSA, FITC-T4, and RLBA datasets, were modelled following the principles proposed by the OECD for the regulatory acceptability of QSARs [53]. A population of about 100 QSAR models was generated for each dataset, as the result of the GA variables subset selection procedure applied in QSARINS. The maximum number of descriptors in each population was kept as low as possible, to reduce the possibility of overfitting, and in agreement to the parsimony principle. The ratio "number of chemicals/number of molecular descriptors", calculated for the best models developed from the three datasets, was about 15, which is largely above the suggested regulatory threshold of five [53]. The best model from each population was selected on the basis of the best balance between fitting and measures of internal cross validation, taking into account both the applicability domain and the residuals distribution of the predicted endpoints. The plot of experimental versus predicted LogRP values from the ANSA, FITC-T4, and RLBA models, are reported in Fig. S1, Fig. S4, and Fig. S7, respectively. All the points are regularly distributed along the diagonal, and no relevant anomalies can be highlighted from these plots.

2.3.1 QSAR model for the ANSA Dataset

The best QSAR model developed for the ANSA dataset is based on four molecular descriptors and is reported in Eq. 2.2:

$$\begin{aligned} \boldsymbol{LogRP_{ANSA}} = & -1.2\ (\pm 0.30) + 1.6 \times 10^{2}\ (\pm 23) \cdot \boldsymbol{AATSC1c} \\ & + 1.7\ (\pm 0.24) \cdot \boldsymbol{PubchemFP381} \\ & + 1.5 \times 10^{-2}\ (\pm 4.8 \times 10^{-3}) \cdot \boldsymbol{ATSC2s} \\ & + 0.28\ (\pm 7.6 \times 10^{-2}) \cdot \boldsymbol{nX} \end{aligned} \tag{2.2}$$

(n° Training set = 59; n° Test set = 20; R^2 = 0.89; $RMSE_{TR}$ = 0.38; Q^2_{loo} = 0.86; Q^2_{lmo} = 0.86; ΔK = 0.088; Q^2_{F3} = 0.88; $RMSE_{TEST}$ = 0.39; R^2_{YS} = 0.070; Q^2_{YS} = −0.11)

In Eq. 2.2, the descriptors are listed in order of relative importance according to their standardized regression coefficients, and the same ordering rule is consistently applied to all QSAR models presented throughout this chapter. Standardized regression coefficients of the descriptors in Eq. 2.2 are reported in Table 2.1. The model fits well (R^2 = 0.89; $RMSE_{TR}$ = 0.38), is internally robust (Q^2_{loo} = 0.86; Q^2_{lmo} = 0.86), and

externally predictive (Q^2_{F3} = 0.88 $RMSE_{TEST}$ = 0.39). Furthermore, R^2_{YS} = 0.070 and Q^2_{YS} = −0.11 exclude chance correlation between the selected descriptors and the response, while positive value of ΔK confirms the absence of multicollinearity. This QSAR covers a wider structural domain compared to literature ones with similar performances and complexity [50].

Fig. S1 shows the plot of experimental versus predicted LogRP values by the ANSA model (Eq. 2.2). As was mentioned above, the analysis of the AD of the model reported in Fig. S2 does not highlight problematic chemicals. Only 2,3,4,5,6-Pentafluorobenzoic acid (ID = 54) has a leverage value (h = 0.26) slightly larger than the threshold (h^* = 0.25). This is probably due to the fact that only ID 54 and ID 73 (3,5-Dibromo-4-hydroxybenzotrifluoride) contain fluorine atoms in Dataset ANSA. Indeed, not surprisingly, ID 73 has a leverage value (h = 0.24) lower but very close to the threshold. However, LogRP is correctly predicted for both chemicals with standardised residuals smaller than ± 2.5 standard deviation units. The QQ plot (Fig. S3) indicates reasonably normally distributed residuals.

The four molecular descriptors selected in Eq. 2.2 are listed and defined in Table S1. These descriptors have a positive relationship with LogRP (i.e., positive value of the regression coefficient in Eq. 2.2), meaning that an increase of their values promotes the binding with hTTR of chemicals. AATSC1c and PubchemFP381 are the most influential descriptors of Eq. 2.2, which are also the two most frequently selected descriptors across the whole population of developed models. AATSC1c is an autocorrelation descriptor which takes into account the spatial distribution of charges along the molecular structure at lag 1, where the lag is the topological distance between pairs of atoms. Previous literature work [32] demonstrated that the dominant binding interactions between halogenated thiophenols and hTTR mainly involve noncovalent interactions, where the spatial distribution of charges can play an important role. Since halogenated thiophenols are those with the highest values of AATSC1c, among chemicals included in the training set, this descriptor could encode information about the tendency of chemicals in forming intermolecular bonds with hTTR. PubchemFP381 is a fragment that, concerning our training set, distinguishes chemicals with a phenoxy group (i.e., PubchemFP381 = 1 when at least one oxygen atom in the structure is bound to an aromatic ring) from those missing this feature (in this case, PubchemFP381 = 0). In the training set, this descriptor allows for the identification of phenols, hydroxylated PBDEs, and hydroxylated PCBs. The selection of this descriptor is in line with literature findings [19,82], showing that the binding with hTTR is favoured by the

presence of hydroxyl groups, aromatic rings, and halogen atoms, which recall the main structural features of THs. This is supported by the fact that most of the training set chemicals with PubchemFP381 = 1 are those with the largest values of experimental RPs. The ATSC2s descriptor reflects how electronic properties, such as polarizability and electronegativity, are distributed along the molecular structure at a lag 2 distance [98]. As seen for AATSC1c, also ATSC2s encodes useful information about the formation of intermolecular bonds with hTTR, such as hydrogen bonds where electronegativity is the relevant factor [99]. Both AATSC1c and ATSC2s are calculated from the Moreau-Broto's autocorrelation coefficient [100]. The nX descriptor is a counter of the number of halogen atoms. Previous studies demonstrated that halogen atoms are relevant to the binding of chemicals with the hTTR [32,35], recalling the presence of iodine atoms in the THs molecular structures. Globally, in the training set, chemicals with the highest experimental response values include from three to six halogen atoms. Chemicals with the lowest LogRP values include a lower number of halogen atoms, ranging from zero to three. After demonstrating the model's predictive potential through external validation, the model was re-developed by combining both training and test sets, in order to use all the available information which is expected to improve the model reliability. The equation of the model is reported below:

$$\begin{aligned} \boldsymbol{LogRP_{ANSA}} \;=\; & -1.1\ (\pm 0.26) \;+\; 1.7 \times 10^{2}\ (\pm 20) \cdot \boldsymbol{AATSC1c} \\ & +\; 1.7\ (\pm 0.20) \cdot \boldsymbol{PubchemFP381} \;+\; 0.29\ (\pm 6.7 \times 10^{-2}) \\ & \cdot \boldsymbol{nX} +\; 1.5 \times 10^{-2}\ (\pm 4.1 \times 10^{-3}) \cdot \boldsymbol{ATSC2s} \end{aligned} \qquad (2.3)$$

(n° Training set = 79; R^2 = 0.89; $RMSE_{TR}$ = 0.38; Q^2_{loo} = 0.87; Q^2_{lmo} = 0.87; ΔK = 0.14; R^2_{YS} = 0.052; Q^2_{YS} = −0.082)

Standardized regression coefficients of the descriptors in Eq. 2.3 are reported in Table 2.1.

Table 2.1. List of descriptors selected in Eq. 2.2 and Eq. 2.3 and their standardized regression coefficients.

Descriptor	**Standardized regression coefficient**	
	Split model (Eq. 2.2)	Full model (Eq. 2.3)
AATSC1c	0.7972	0.7724
PubchemFP381	0.7034	0.7238
ATSC2s	0.3645	0.3508
nX	0.3626	0.3607

As expected, taking into account the good results of the external validation of Eq. 2.2, as well as the distribution of the training and test sets in Fig. S1 and Fig. S2, the values of the intercept and of the coefficients in Eq. 2.2 and in Eq. 2.3 are very similar. This further confirms the robustness of the selected descriptors, and Eq. 2.3 can be suggested as a reliable QSAR for the prediction of $LogRP_{ANSA}$ of new chemicals.

2.3.2 QSAR model for the FITC-T4 Dataset

The equation of the best QSAR model developed using the information included in the Dataset FITC-T4 is:

$$\begin{aligned} \boldsymbol{LogRP_{FITC-T4}} &= -0.84\ (\pm 0.35)\ +\ 0.32\ (\pm 4.8 \times 10^{-2}) \cdot \boldsymbol{naasC} \\ &-\ 1.6\ (\pm 0.48) \cdot \boldsymbol{SpMin4\_Bhs} \\ &+\ 1.0 \times 10^{-2}\ (\pm 6.2 \times 10^{-3}) \cdot \boldsymbol{VE3\_Dzs} \end{aligned} \tag{2.4}$$

(n° Training set = 38; n° Test set = 12; R^2 = 0.85; $RMSE_{TR}$ = 0.38; Q^2_{loo} = 0.81; Q^2_{lmo} = 0.80; ΔK = 0.030; Q^2_{F3} = 0.83; $RMSE_{TEST}$ = 0.40; R^2_{YS} = 0.082; Q^2_{YS} = −0.15)

Standardized regression coefficients of the descriptors in Eq. 2.4 are reported in Table 2.2. This model has good fitting (R^2 = 0.85; $RMSE_{TR}$ = 0.38), is internally robust (Q^2_{loo} = 0.81; Q^2_{lmo} = 0.80), and has a good external predictive potential (Q^2_{F3} = 0.83; $RMSE_{TEST}$ = 0.40). The absence of chance correlation between the descriptors and the response is confirmed by the Y-scrambling procedure (R^2_{YS} = 0.082; Q^2_{YS} = −0.15), while the absence of multicollinearity is confirmed by the positive value of ΔK.

The plot of experimental versus predicted LogRP values by the FITC-T4 model is reported in Fig. S4.

The Williams plot, reported in Fig. S5, shows that all the chemicals fall into the structural applicability domain of the model. Only tetrabromobisphenol A-mono (allyl ether) (ID = 21) has a standardised residual in prediction of –2.6, slightly exceeding the reference range of ± 2.5 standard deviation units. It is interesting to note that this chemical is the only bisphenol derivative in the test set, having also the largest value of experimental RP compared to all the bisphenol derivatives included in the dataset.

This may be the reason of the large residual in prediction. The QQ plot (Fig. S6) shows reasonably normally distributed residuals. The three molecular descriptors selected in Eq. 2.4 are listed in Table S2.

The most important molecular descriptor is naasC, positively correlated with the response. This suggests that larger values of this descriptor contribute to a greater binding ability with hTTR. NaasC is an electrotopological state index [101], encoding for electronic and topological information, such as electronegativity and polarizability, as seen for ATSC2s in Eq. 2.2; it counts the number of bonds that involve aromatic carbons, excluding those with hydrogen and carbon atoms in the same aromatic ring. This descriptor, in the training set of the FITC-T4 model, discriminates between more and less substituted diphenyls (mainly in terms of hydroxyl groups, halogen atoms, ether bonds), phenols, and not aromatic chemicals (such as PFAS), with naasC values decreasing in this order. Globally, chemicals with the highest values of experimental RPs are those with the highest values of naasC. Overall, it seems that the presence of hydroxyl groups, aromatic rings, and halogen atoms, is important in promoting the binding affinity with the hTTR. These results are in agreement with those highlighted for the ANSA model, despite the use of different training sets for the development of the two models.

The second most important molecular descriptor is SpMin4_Bhs, which takes into account the topology of the chemicals and has negative sign in Eq. 2.4. Indeed, in this dataset, the smallest values of this descriptor are calculated for PFAS, which are known to bind with hTTR [30,47,62,74]. The correlation between SpMin4_Bhs and LogRP values for the 14 PFAS in the training set is −0.88. These results are supported by the dependency of SpMin4_Bhs with the carbon chain length of PFAS, which specularly reflects the dependency of the binding capacity of PFAS with their own carbon chain length. In the literature, it is demonstrated that PFAS with a terminal carboxylic acid group, and with a carbon chain length between six and ten, are more potent hTTR binders than the equivalents with a longer or shorter carbon chain length [47]. For sulfonic PFAS, a carbon chain length equal to eight optimises the binding with hTTR [47,74]. In the studied dataset, SpMin4_Bhs values are lower for PFAS with a terminal carboxylic acid group and with a carbon chain length between six and ten, while are higher for the equivalents with a longer or shorter carbon chain length. Similarly, SpMin4_Bhs values exhibited a decreasing trend with increasing length of the carbon chain in sulfonic PFAS, reaching a minimum value for the eight-carbon sulfonic PFAS (sulfonic PFAS with a carbon chain length larger than eight are missing in the studied

dataset). Therefore, the selection of this descriptor in Eq. 2.4, seems to be particularly related to PFAS disruptive effect on hTTR functions, considering its negative sign in Eq. 2.4. It should be noted that naasC and SpMin4_Bhs are two of the most frequently selected descriptors across the population of developed models, and the two most frequent ones in the population of models developed with three variables. VE3_Dzs is the less influential descriptor, which is calculated from the Barysz distance matrix that accounts simultaneously for heteroatoms and multiple bonds. In Eq. 2.4, it has a positive correlation with the response, suggesting that heteroatoms and multiple bonds together have a role in increasing the binding affinity with hTTR. In particular, the value of this descriptor increases with the response for chemicals that share the same value of naasC. Therefore, VE3_Dzs provides additional information to discriminate different molecular and atomic aspects for chemicals having the same value of naasC.

Finally, FITC-T4 model was re-developed by combining both training and test sets in order to use all the available information. The equation of the full model is reported as follows (Eq. 2.5):

$$\begin{aligned} \boldsymbol{LogRP_{FITC-T4}} &= -0.97\ (\pm 0.30)\ +\ 0.33\ (\pm 4.3 \times 10^{-2}) \cdot \boldsymbol{naasC} \\ &-\ 1.5\ (\pm 0.42) \cdot \boldsymbol{SpMin4\_Bhs} \\ &+\ 1.1 \times 10^{-2}\ (\pm 5.8 \times 10^{-3}) \cdot \boldsymbol{VE3\_Dzs} \end{aligned} \tag{2.5}$$

(n° Training set = 50; R^2 = 0.84; $RMSE_{TR}$ = 0.38; Q^2_{loo} = 0.81; Q^2_{lmo} = 0.80; ΔK = 0.058; R^2_{YS} = 0.061; Q^2_{YS} = −0.11)

Standardized regression coefficients of the descriptors in Eq. 2.5 are reported in Table 2.2.

Table 2.2. List of descriptors selected in Eq. 2.4 and Eq. 2.5 and their standardized regression coefficients.

Descriptor	**Standardized regression coefficient**	
	Split model (Eq. 2.4)	Full model (Eq. 2.5)
naasC	1.1369	1.125
SpMin4_Bhs	− 0.579	− 0.5416
VE3_Dzs	0.2307	0.2195

As was described for Eq. 2.2 and Eq. 2.3, the values of the intercept and of the coefficients in the Eq. 2.4 and Eq. 2.5 are very similar. Eq. 2.5 is the final QSAR for the prediction of $LogRP_{FITC\text{-}T4}$ for new chemicals.

2.3.3 QSAR model for the RLBA Dataset

The first modelling attempt performed on Dataset RLBA highlighted four chemicals as frequent outliers in the population of models, which may negatively affect their performances: 3,3',5,5'-tetrachlorobiphenyl (ID = 46), 3,3',4,4',5-pentachlorobiphenyl (ID = 93), 3,3',4,5,5'-pentachlorobiphenyl (ID = 75), and perfluorobutanesulfonic acid (ID = 109). The LogRP value (0.85) of 3,3',5,5'-tetrachlorobiphenyl is 2 log units larger than LogRP reported for the other four tetrachlorobiphenyls, which ranges from −1.1 to −1.3. By excluding the two pentachlorobiphenyls listed above from the dataset, LogRP ranges from −1.8 to 0.44 for the other nine pentachlorobiphenyls, while 3,3',4,4',5-pentachlorobiphenyl and 3,3',4,5,5'-pentachlorobiphenyl LogRP values are respectively −3.4 and 0.91. Perfluorobutanesulfonic acid is also an outlier, whose LogRP value (−2.5) is smaller compared to the other three perfluorinated sulfonic acids included in the dataset (LogRP range between −1.2 and −0.77). It is important to highlight that the LogRP values of these outliers are not averages of experimental values taken from different sources, which internal variability may have affected the quality of the model.

By removing the aforementioned four chemicals from the dataset, a population of models was developed, and the following QSAR was selected:

$$\begin{aligned} \boldsymbol{LogRP_{RLBA}} = & -32\ (\pm 7.0)\ +\ 1.8\ (\pm 0.29) \cdot \boldsymbol{PubchemFP590} \\ & +\ 7.7\ (\pm 1.8) \cdot \boldsymbol{SpMax1\_Bhe}\ -\ 1.3\ (\pm 0.43) \\ & \cdot \boldsymbol{PubchemFP18}\ -\ 1.2\ (\pm 0.39) \cdot \boldsymbol{GATS5c}\ -\ 20\ (\pm 7.9) \\ & \cdot \boldsymbol{AATSC1e}\ +\ 4.9 \times 10^{-3}\ (\pm 2.0 \times 10^{-3}) \cdot \boldsymbol{AATS4v} \end{aligned} \tag{2.6}$$

(n° Training set = 100; n° Test set = 33; R^2 = 0.81; $RMSE_{TR}$ = 0.52; Q^2_{loo} = 0.77; Q^2_{lmo} = 0.77; ΔK = 0.073; Q^2_{F3} = 0.69; $RMSE_{TEST}$ = 0.66; R^2_{YS} = 0.061; Q^2_{YS} = −0.088)

Standardized regression coefficients of the descriptors in Eq. 2.6 are reported in Table 2.3. The RLBA model has good validation metrics values, slightly worse compared to the ANSA and FITC-T4 models. This could be due to the higher structural

heterogeneity and size of the RLBA dataset, compared to the other datasets. However, the model has good fitting according to R^2 and $RMSE_{TR}$ values, Q^2_{loo} and Q^2_{lmo} values support the model internal robustness, while Q^2_{F3} and $RMSE_{TEST}$ support its external predictivity when applied to new chemicals. The R^2_{YS} and Q^2_{YS} values exclude chance correlation, while ΔK value greater than zero excludes multicollinearity. The removal of the aforementioned outliers led to an improvement of the model performances in terms of fitting and internal robustness (i.e. former QSARs with six descriptors had R^2 from 0.72 to 0.75, $RMSE_{TR}$ from 0.61 to 0.64, while Q^2_{loo} from 0.68 to 0.71).

Experimental versus predicted LogRP values of RLBA model can be found in Fig. S7.

The Williams plot, reported in Fig. S8, shows that most chemicals are within the applicability domain of the model, confirming the absence of both response and structural outliers, with an exception for the chemical perfluoro-n-pentanoic acid (ID = 120). Despite this chemical has a leverage value (h = 0.24) greater than the defined threshold (h^* = 0.21), the model predicts reasonably LogRP (standardised residual smaller than ±2.5 standard deviation units). Residuals are reasonably normally distributed according to the QQ plot (Fig. S9). The selected molecular descriptors are listed in Table S3.

The PubchemFP590 descriptor encodes for the presence of hydroxylated aromatic rings in the molecular structure (value 1 of the descriptor). It is the most relevant variable in Eq. 2.6, and its positive sign indicates that the presence of the molecular fragment leads to an increase of the binding affinity with hTTR. In the training set, this descriptor identifies chemicals such as hydroxylated PCBs and PBDEs, phenols, parabens, hydroxylated dioxins and furans, hydroxylated xanthones, and bisphenol A derivatives. Most of the chemicals with PubchemFP590 = 1, included in the training set, are those with the greatest values of experimental RPs. This is in agreement with the structural information selected in ANSA and FITC-T4 models, despite using different training sets. Furthermore, PubchemFP590 highlights the same critical structural feature that promotes the binding with hTTR, previously highlighted.

The second most relevant descriptor in Eq. 2.6 is SpMax1_Bhe, which is calculated from the Burden matrix and is weighted by relative Sanderson electronegativities. SpMax1_Bhe belongs to the same family of descriptors as SpMin4_Bhs, selected in the FITC-T4 model. In Eq. 2.6, this descriptor has a positive sign, indicating its correlation with the response. Chemicals with the greatest values of this descriptor are mainly PCBs, hydroxylated PCBs, hydroxylated xanthones, then PBDEs, hydroxylated PBDEs, PFAS, bisphenol A derivatives, and finally phenols. The role of this descriptor

may be to take into account the binding ability of non-hydroxylated biphenyls (i.e. PCBs, PBDEs) and PFAS, which is not considered by PubchemFP590. Phenols are the compounds with the smallest SpMax1_Bhe value and the lowest binding affinity, compared to chemicals with two aromatic rings. This is not surprising, since the binding activity of a chemical is affected by the structural similarity to THs (which is high for halogenated and hydroxylated biphenyls and biphenyls ethers).

The third most relevant molecular descriptor is PubchemFP18, which has an absolute standardized coefficient comparable to SpMax1_Bhe in Eq. 2.6 (–0.45 and 0.46, respectively). In this training set, PubchemFP18 identifies chemicals with one or more oxygen atoms (value 1 of the descriptor) but no hydroxyl groups bound to the aromatic ring. Therefore, the combination of PubchemFP18 and PubchemFP590 allows for the correct modelling and the discrimination across the activity of hydroxylated aromatics, mentioned above, from non-hydroxylated aromatics (mainly PBDEs) and not aromatic but hydroxylated compounds (mainly carboxylic and sulfonic PFAS).

The three less important molecular descriptors selected in Eq. 2.6 are GATS5c, AATSC1e, and AATS4v, which belong to the 2D autocorrelation descriptors, which are known for their usefulness and applicability [99]. In this context, these descriptors capture the topological distribution of several properties of the chemical structures. Differently from the other autocorrelation descriptors, which are calculated from the Moreau-Broto's autocorrelation coefficient, GATS5c is calculated from the Geary's coefficient and is related to the distribution of charges across the molecules.

Once the model's predictive ability was demonstrated through external validation, training and test sets were pooled and the model was re-developed, resulting in the following equation:

$$\begin{aligned} \boldsymbol{LogRP_{RLBA}} = & -30\ (\pm 6.4) + 1.9\ (\pm 0.28) \cdot \boldsymbol{PubchemFP590} \\ & - 1.5\ (\pm 0.42) \cdot \boldsymbol{PubchemFP18} + 7.4\ (\pm 1.6) \\ & \cdot \boldsymbol{SpMax1\_Bhe} - 22\ (\pm 7.1) \cdot \boldsymbol{AATSC1e} - 1.3\ (\pm 0.38) \\ & \cdot \boldsymbol{GATS5c} + 4.9 \times 10^{-3}\ (\pm 1.8 \times 10^{-3}) \cdot \boldsymbol{AATS4v} \end{aligned} \tag{2.7}$$

(n° Training set = 133; R^2 = 0.78; Q^2_{loo} = 0.75; Q^2_{lmo} = 0.75; $RMSE_{TR}$ = 0.55; ΔK = 0.067; R^2_{YS} = 0.045; Q^2_{YS} = −0.065)

Standardized regression coefficients of the descriptors in Eq. 2.7 are reported in Table 2.3.

Table 2.3. List of descriptors selected in Eq. 2.6 and Eq. 2.7 and their standardized regression coefficients.

Descriptor	**Standardized regression coefficient**	
	Split model (Eq. 2.6)	Full model (Eq. 2.7)
PubchemFP590	0.7523	0.7815
SpMax1_Bhe	0.458	0.4482
PubchemFP18	− 0.449	− 0.5259
GATS5c	− 0.2913	− 0.3057
AATSC1e	− 0.2859	− 0.3545
AATS4v	0.2478	0.2493

Consistently with former models, considering the similarities between the Eq. 2.6 and Eq. 2.7, Eq. 2.7 is suggested for the prediction of $LogRP_{RLBA}$ of new chemicals.

Finally, it is interesting to highlight that the main structural features selected in Eq. 2.2, Eq. 2.4, and Eq. 2.6, are similar. Although the structural information included in the training set of each model led to the selection of different sets of molecular descriptors, the mechanistic interpretation of the overall set of the selected descriptors mainly converges towards the same key molecular structures that affect binding capacity of compounds, as was previously described.

PubchemFP381, selected in the ANSA model, distinguishes chemicals with a phenoxy group (mainly hydroxylated aromatic compounds). Similar structural information (i.e. the discrimination of non-hydroxylated aromatic compounds from hydroxylated aromatic compounds (e.g., PBDEs and PCBs from their hydroxylated metabolites)) is encoded by the combination of PubchemFP590 and PubchemFP18 selected in the RLBA model. Furthermore, the relevance of substituted aromatic rings (in terms of hydroxyl groups and halogen atoms, based on our training set) is encoded by naasC descriptor selected in the FITC-T4 model.

The descriptor nX selected in the ANSA model counts the number of halogen atoms, as well as VE3_Dzs, selected in the FITC-T4 model, is sensitive to this structural feature.

Autocorrelation descriptors, selected in the ANSA and RLBA models, encode for the topological distribution of different properties along the chemical structure, which can influence the tendency of compounds in forming chemical bonds with hTTR. In particular, the same descriptor AATSC1, weighted on charges (i.e., AATSC1c) and on

electronegativity (i.e., AATSC1e), is selected in the ANSA and RLBA models, respectively. The influence of charges is also encoded by the descriptor GATS5c, which is selected in the RLBA model.

2.3.4 Models application to Dataset II and PCA of Dataset III

Equations calibrated on the full datasets ANSA, FITC-T4, and RLBA, respectively Eq. 2.3, Eq. 2.5 and Eq. 2.7, were applied to predict missing experimental values of T4-hTTR competing potencies of 223 chemicals included in Dataset I, leading to Dataset II. Among these chemicals, 73 belonging to different chemical classes were discarded from further analysis because of unreliable predictions provided by at least one QSAR (see Section 2.2.1), mainly due to poor representation of these chemicals in the QSARs training sets. In particular, predictions were excluded for PFAS and non-hydroxylated PBDEs (ANSA model), as well as for phosphates and fatty acids (FITC-T4 and RLBA models). Experimental and reliable LogRP predictions were available for 150 chemicals and included in Dataset III, which was explored by PCA. Fig. 2.1 shows the PCA biplot (scores and loadings) of the first two components, where the first component explains more than 65 % of the total variance (cumulative explained variance for PC1 and PC2 is about 85 % of the total variance). Scores and loadings values calculated for the first two principal components are reported in Supplementary Material S2.

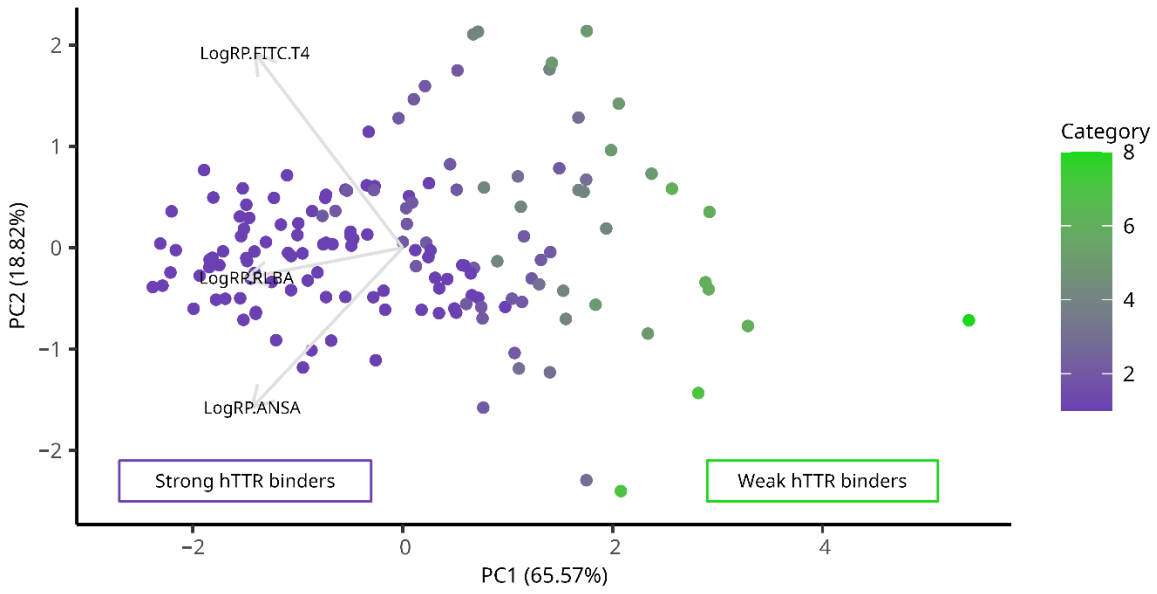

Figure 2.1. Biplot resulting from PCA analysis carried out on Dataset III to compare experimental or predicted LogRPs for the three in vitro assays ANSA, FITC–T4, and RLBA. For graphical visualization, all the compounds in Dataset III have been categorised on the basis of their respective experimental or

predicted LogRPs, according to the classes defined by Yang et al. (strong hTTR binder: LogRP ≥ −1.26; moderate hTTR binder: −2.26 ≤ LogRP < −1.26; weak hTTR binder: LogRP < −2.26) [50]. Different combinations of the aforementioned classes in experimental or predicted data define categories from 1 to 8 (e.g., a chemical falls into category 1 if it is a strong hTTR binder according to each experimental or predicted LogRP; a chemical falls into category 8 if it is a weak hTTR binder according to each experimental or predicted LogRP). Additional details on the definition of the eight categories are reported in Supplementary Material S2.

The loadings are positively correlated and equally important along PC1 (internal correlations: LogRP ANSA – LogRP FITC-T4 = 0.44; LogRP ANSA – LogRP RLBA = 0.52; LogRP FITC-T4 – LogRP RLBA = 0.49), while PC2 highlights intra methods differences between LogRP FITC-T4 and LogRP ANSA (loadings are opposite in sign).

Considering the values of the loadings and their agreement in the biplot, PC1 ranks chemicals from right to left according to their increasing T4-hTTR competing potency. Chemicals on the left side of the graph are strong hTTR binders with large LogRP values, either experimental or predicted by the three QSARs. On the contrary, chemicals on the right side of the graph have small values of LogRP, and consequently are weak hTTR binders. As expected, the strongest hTTR binders have two hydroxylated or non-hydroxylated aromatic rings with a high number of halogen atoms (at least three), such as PCBs, PBDEs, and their hydroxylated metabolites, halogenated bisphenol A derivatives (e.g. tetrabromobisphenol A), halogenated dioxins (e.g. 2-hydroxy-1,3,7,8-tetrachlorodibenzo-4-dioxin), highly halogenated phenols (e.g. pentabromophenol), and thiophenols (e.g. pentachlorothiophenol). Chemicals with less or no substituents as hydroxyl groups and halogen atoms, with one or two aromatic rings, and with heteroatoms such as nitrogen and sulphur in their structure, are less powerful competitors with T4 in hTTR binding. Examples of weaker hTTR binders are non-hydroxylated and low halogenated PCBs and PBDEs, and related sulfated metabolites, one or two halogen-substituted thiophenols (e.g. 2-bromothiophenol) and phenols (e.g. 3-chlorophenol), xanthone derivatives (e.g. 2-hydroxy-1-methoxyxanthone), and parabens (e.g. ethyl- and propylparaben). Tris (2-chloroethyl) phosphate has the lowest LogRP values experimentally measured with the ANSA method and predicted by FITC-T4 and RLBA models (all LogRPs less than –3). This chemical does not have any aromatic ring.

Finally, PC2, which explains about 20 % of the residual variance, highlights that intra methods differences are more evident between LogRP-FITC and LogRP-ANSA values,

in particular for moderate/weak binders. These differences may be explained by a prevalence of predicted LogRP values by Eq. 2.3 and/or Eq. 2.5 for substances with PC2 scores smaller than −1 or PC2 scores greater than 1. Therefore, the separation of the loadings LogRP FITC-T4 and LogRP-ANSA along PC2, is probably due to the uncertainty in prediction and to the intra-methods experimental uncertainty.

2.4 Conclusions

On the basis of the increasing concern against EDCs and the need for their early identification using *in silico* approaches, this work was primarily focused on the development of three new MLR QSARs, to detect THSDCs by predicting the competition with the T4 for the binding with hTTR. These QSARs were developed using new and curated datasets which included, to the best of our knowledge, all the currently available quantitative experimental binding affinity values measured using three *in vitro* assays (i.e., ANSA, FITC-T4, and RLBA). QSARs developed to predict LogRP values for the ANSA and RLBA datasets, which are based on heterogeneous training sets characterised by broad response ranges (from weak to strong hTTR binders), have wider applicability domains compared to existing similar QSARs. To our knowledge, no QSARs have been developed, so far, to predict LogRP measured with the FITC-T4 methodology, other than the here proposed model.

The mechanistic interpretation of the descriptors selected in Eq. 2.2, Eq. 2.4 and Eq. 2.6, is coherent with the most important structural features which are known to be involved in the binding of chemicals to hTTR, such as hydroxylated and halogenated aromatic rings, as well as descriptors encoding for specific electronic environments. The binding of chemicals to hTTR was further investigated running PCA analysis on experimental and reliable LogRPs predictions for the three datasets. PCA ranked the studied chemicals along PC1, according to their relative competitive potency and confirmed that the structural features mentioned above are relevant to distinguish between strong and weak binding activity.

A limitation of this study is related to the fact that some specific chemical classes (i.e. PFAS, non-hydroxylated PBDEs, phosphates, and fatty acids) were not included in all the training sets, and therefore they were predicted outside the applicability domain of one or more of the here proposed QSARs. This led to the exclusion of these chemicals from the PCA analysis, which underlines the need for additional *in vitro* tests in these specific areas of the structural space. These new data would be useful to refine the existing QSARs, and to generate new models covering wider structural domains. Furthermore, as the proposed models target only hTTR binders, they may overestimate the binding affinity of inactive compounds.

However, the here proposed QSARs, which are characterised by low complexity and are statistically robust and predictive, can be applied for the early screening of THSDCs on the basis of the molecular structure. To ease their application and support

hazard assessment procedures, they have been implemented in the QSAR-ME Profiler beta version 1.02 software (freely available at https://dunant.dista.uninsubria.it/qsar/).

Acknowledgments and Funding: This work was supported by the PhD Program in Chemical and Environmental Sciences (DiSCA, University of Insubria) (PhD scholarship to Marco Evangelista).

Supporting information: Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jhazmat.2024.136188.

2.5 References

[1] De Coster, S., van Larebeke, N., 2012. J. Environ. Public Health 2012, 713696, https://doi.org/10.1155/2012/713696.

[2] Kahn, L.G., Philippat, C., Nakayama, S.F., Slama, R., Trasande, L., 2020. Lancet Diabetes Endocrinol. 8, 703–718, https://doi.org/10.1016/S2213-8587(20)30129-7.

[3] World Health Organisation (WHO), 2002. Damstra, T., Barlow, S., Bergman, A., Kavlock, R., Van Der Kraak, G., editors. Global assessment of the state-of-the-science of endocrine disruptors. Geneva: World Health Organization, https://www.who.int/publications/i/item/WHO-PSC-EDC-02.2 (last accessed on 06 October 2025).

[4] Khan, S., Naushad, Mu., Govarthanan, M., Iqbal, J., Alfadul, S.M., 2022. Environ. Res. 207, 112609, https://doi.org/10.1016/j.envres.2021.112609.

[5] Zoeller, R.T., Brown, T.R., Doan, L.L., Gore, A.C., Skakkebaek, N.E., Soto, A.M., Woodruff, T.J., Vom Saal, F.S., 2012. Endocrinology 153, 4097–4110, https://doi.org/10.1210/en.2012-1422.

[6] Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocr. Rev. 30, 293–342, https://doi.org/10.1210/er.2009-0002.

[7] European Commission, 1999. Community Strategy for Endocrine Disrupters - a range of substances suspected of interfering with the hormone systems of humans and wildlife. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:1999:0706:FIN:EN:PDF (last accessed on 06 October 2025).

[8] Andersson, N., Arena, M., Auteri, D., Barmaz, S., Grignard, E., Kienzler, A., Lepper, P., Lostia, A.M., Munn, S., Parra Morte, J.M., Pellizzato, F., Tarazona, J., Terron, A., Van der Linden, S., 2018. EFSA J. 16, e05311, https://doi.org/10.2903/j.efsa.2018.5311.

[9] European Commission, 2019. The European Green Deal. https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (last accessed on 06 October 2025).

[10] European Commission, 2020. Chemicals Strategy for Sustainability towards a Toxic-free Environment. https://environment.ec.europa.eu/strategy/chemicals-strategy_en (last accessed on 06 October 2025).

[11] European Commission, 2022. Commission Delegated Regulation (EU) 2023/707 of 19 December 2022 amending Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures (Text with EEA relevance).

[12] Calsolaro, V., Pasqualetti, G., Niccolai, F., Caraccio, N., Monzani, F., 2017. Int. J. Mol. Sci. 18, 2583, https://doi.org/10.3390/ijms18122583.

[13] Crofton, K.M., 2008. Int. J. Androl. 31, 209–223 https://doi.org/10.1111/j.1365-2605.2007.00857.x

[14] Köhrle, J., Frädrich, C., 2021. Best Pract. Res. Clin. Endocrinol. Metab. 35, 101562, https://doi.org/10.1016/j.beem.2021.101562.

[15] Sokal, A., Jarmakiewicz-Czaja, S., Tabarkiewicz, J., Filip, R., 2021. Nutrients 13, 867, https://doi.org/10.3390/nu13030867.

[16] Miller, M.D., Crofton, K.M., Rice, D.C., Zoeller, R.T., 2009. Environ. Health Perspect. 117, 1033–1041, https://doi.org/10.1289/ehp.0800247.

[17] Noyes, P.D., Friedman, K.P., Browne, P., Haselman, J.T., Gilbert, M.E., Hornung, M.W., Barone, S., Crofton, K.M., Laws, S.C., Stoker, T.E., Simmons, S.O., Tietge, J.E., Degitz, S.J., 2019. Environ. Health Perspect. 127, 095001, https://doi.org/10.1289/EHP5297.

[18] Boas, M., Feldt-Rasmussen, U., Main, K.M., 2012. Mol. Cell. Endocrinol. 355, 240–248, https://doi.org/10.1016/j.mce.2011.09.005.

[19] Brucker-Davis, F., 1998. Thyroid 8, 827–856 https://doi.org/10.1089/thy.1998.8.827.

[20] Bernasconi, C., Langezaal, I., Bartnicka, J., Asturiol, D., Bowe, G., Coecke, S., Kienzler, A., Liska, R., Milcamps, A., Munoz-Pineiro, M.A., Pistollato, F., Whelan, M., 2023. Publications Office of the European Union, Luxembourg, https://doi.org/10.2760/862948.

[21] Kortenkamp, A., Axelstad, M., Baig, A.H., Bergman, Å., Bornehag, C.-G., Cenijn, P., Christiansen, S., Demeneix, B., Derakhshan, A., Fini, J.-B., Frädrich, C., Hamers, T., Hellwig, L., Köhrle, J., Korevaar, T.I.M., Lindberg, J., Martin, O., Meima, M.E., Mergenthaler, P., Nikolov, N., Du Pasquier, D., Peeters, R.P., Platzack, B., Ramhøj, L., Remaud, S., Renko, K., Scholze, M., Stachelscheid, H., Svingen, T., Wagenaars, F., Wedebye, E.B., Zoeller, R.T., 2020. Int. J. Mol. Sci. 21, 3123, https://doi.org/10.3390/ijms21093123.

[22] Organisation for Economic Co-operation and Development (OECD), 2014. New Scoping Document on in vitro and ex vivo Assays for the Identification of Modulators of Thyroid Hormone Signalling. https://doi.org/10.1787/9789264274716-en.

[23] Giammanco, M., Di Liegro, C.M., Schiera, G., Di Liegro, I., 2020. Int. J. Mol. Sci. 21, 4140, https://doi.org/10.3390/ijms21114140.

[24] Schroeder, A.C., Privalsky, M.L., 2014. Front. Endocrinol. 5, 40, https://doi.org/10.3389/fendo.2014.00040.

[25] Richardson, S.J., 2008. Mol. Cell. Endocrinol. 293, 32–42, https://doi.org/10.1016/j.mce.2008.04.002.

[26] Alshehri, B., D'Souza, D.G., Lee, J.Y., Petratos, S., Richardson, S.J., 2015. J. Neuroendocrinol. 27, 303–323, https://doi.org/10.1111/jne.12271.

[27] Gião, T., Saavedra, J., Cotrina, E., Quintana, J., Llop, J., Arsequell, G., Cardoso, I., 2020. Int. J. Mol. Sci. 21, 2075, https://doi.org/10.3390/ijms21062075.

[28] Schussler, G.C., 2000. Thyroid 10, 141–149, https://doi.org/10.1089/thy.2000.10.141.

[29] Meerts, I.A.T.M., Assink, Y., Cenijn, P.H., van den Berg, J.H.J., Weijers, B.M., Bergman, Å., Koeman, J.H., Brouwer, A., 2002. Toxicol. Sci. 68, 361–371, https://doi.org/10.1093/toxsci/68.2.361.

[30] Weiss, J.M., Andersson, P.L., Lamoree, M.H., Leonards, P.E.G., van Leeuwen, S.P.J., Hamers, T., 2009. Toxicol. Sci. 109, 206–216, https://doi.org/10.1093/toxsci/kfp055.

[31] Šauer, P., Švecová, H., Grabicová, K., Gönül Aydın, F., Mackuľak, T., Kodeš, V., Blytt, L.D., Henninge, L.B., Grabic, R., Kocour Kroupová, H., 2021. Sci. Total Environ. 751, 141801, https://doi.org/10.1016/j.scitotenv.2020.141801.

[32] Yang, X., Ou, W., Zhao, S., Wang, L., Chen, J., Kusko, R., Hong, H., Liu, H., 2021. Chemosphere 280, 130627, https://doi.org/10.1016/j.chemosphere.2021.130627.

[33] Hamers, T., Kamstra, J.H., Cenijn, P.H., Pencikova, K., Palkova, L., Simeckova, P., Vondracek, J., Andersson, P.L., Stenberg, M., Machala, M., 2011. Toxicol. Sci. 121, 88–100, https://doi.org/10.1093/toxsci/kfr043.

[34] Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Kester, M.H.A., Andersson, P.L., Legler, J., Brouwer, A., 2006. Toxicol. Sci. 92, 157–173, https://doi.org/10.1093/toxsci/kfj187.

[35] Lans, M.C., Klasson-Wehler, E., Willemsen, M., Meussen, E., Safe, S., Brouwer, A., 1993. Chem. Biol. Interact. 88, 7–21, https://doi.org/10.1016/0009-2797(93)90081-9.

[36] Ren, X.M., Guo, L.-H., 2012. Environ. Sci. Technol. 46, 4633–4640, https://doi.org/10.1021/es2046074.

[37] Ramhøj, L., Axelstad, M., Baert, Y., Cañas-Portilla, A.I., Chalmel, F., Dahmen, L., De La Vieja, A., Evrard, B., Haigis, A.-C., Hamers, T., Heikamp, K., Holbech, H., Iglesias-Hernandez, P., Knapen, D., Marchandise, L., Morthorst, J.E., Nikolov, N.G., Nissen, A.C.V.E., Oelgeschlaeger, M., Renko, K., Rogiers, V., Schüürmann, G., Stinckens, E., Stub, M.H., Torres-Ruiz, M., Van Duursen, M., Vanhaecke, T., Vergauwen, L., Wedebye, E.B., Svingen, T., , 2023. Front. Toxicol. 5, 1189303, https://doi.org/10.3389/ftox.2023.1189303.

[38] Organisation for Economic Co-operation and Development (OECD), 2009. Test No. 455: The Stably Transfected Human Estrogen Receptor-alpha Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals. https://doi.org/10.1787/9789264076372-en.

[39] Organisation for Economic Co-operation and Development (OECD), 2023. Test No. 458: Stably Transfected Human Androgen Receptor Transcriptional Activation Assay for Detection of Androgenic Agonist and Antagonist Activity of Chemicals. https://doi.org/10.1787/9789264264366-en.

[40] Eytcheson, S.A., Olker, J.H., Friedman, K.P., Hornung, M.W., Degitz, S.J., 2023. Regul. Toxicol. Pharmacol. 144, 105491, https://doi.org/10.1016/j.yrtph.2023.105491.

[41] Mitchell, C.A., Burden, N., Bonnell, M., Hecker, M., Hutchinson, T.H., Jagla, M., LaLone, C.A., Lagadic, L., Lynn, S.G., Shore, B., Song, Y., Vliet, S.M., Wheeler, J.R., Embry, M.R., 2023. Environ. Toxicol. Chem. 42, 757–777, https://doi.org/10.1002/etc.5584.

[42] Schneider, M., Pons, J.-L., Labesse, G., Bourguet, W., 2019. Endocrinology 160, 2709–2716, https://doi.org/10.1210/en.2019-00382.

[43] Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., Zakharov, A., Worth, A., Richard, A.M., Grulke, C.M., Trisciuzzi, D., Fourches, D., Horvath, D., Benfenati, E., Muratov, E., Wedebye, E.B., Grisoni, F., Mangiatordi, G.F., Incisivo, G.M., Hong, H., Ng, H.W., Tetko, I.V., Balabin, I., Kancherla, J., Shen, J., Burton, J., Nicklaus, M., Cassotti, M., Nikolov, N.G., Nicolotti, O., Andersson, P.L., Zang, Q., Politi, R., Beger, R.D., Todeschini, R., Huang, R., Farag, S., Rosenberg, S.A., Slavov, S., Hu, X., Judson, R.S., 2016. Environ. Health Perspect. 124, 1023–1033, https://doi.org/10.1289/ehp.1510267.

[44] Mansouri, K., Kleinstreuer, N., Abdelaziz, A.M., Alberga, D., Alves, V.M., Andersson, P.L., Andrade, C.H., Bai, F., Balabin, I., Ballabio, D., Benfenati, E., Bhhatarai, B., Boyer, S., Chen, J., Consonni, V., Farag, S., Fourches, D., García-Sosa, A.T., Gramatica, P., Grisoni, F., Grulke, C.M., Hong, H., Horvath, D., Hu, X., Huang, R., Jeliazkova, N., Li, J., Li, X., Liu, H., Manganelli, S., Mangiatordi, G.F., Maran, U., Marcou, G., Martin, T., Muratov, E., Nguyen, D.-T., Nicolotti, O., Nikolov, N.G., Norinder, U., Papa, E., Petitjean, M., Piir, G., Pogodin, P., Poroikov, V., Qiao, X., Richard, A.M., Roncaglioni, A., Ruiz, P., Rupakheti, C., Sakkiah, S., Sangion, A., Schramm, K.-W., Selvaraj, C., Shah, I., Sild, S., Sun, L., Taboureau, O., Tang, Y., Tetko, I.V., Todeschini, R., Tong, W., Trisciuzzi, D., Tropsha, A., Van Den Driessche, G., Varnek, A., Wang, Z., Wedebye, E.B., Williams, A.J., Xie, H., Zakharov, A.V., Zheng, Z., Judson, R.S., 2020. Environ. Health Perspect. 128, 027002, https://doi.org/10.1289/EHP5580.

[45] Richard, A.M., Huang, R., Waidyanatha, S., Shinn, P., Collins, B.J., Thillainadarajah, I., Grulke, C.M., Williams, A.J., Lougee, R.R., Judson, R.S., Houck, K.A., Shobair, M., Yang, C., Rathman, J.F., Yasgar, A., Fitzpatrick, S.C., Simeonov, A., Thomas, R.S., Crofton, K.M., Paules, R.S., Bucher, J.R., Austin, C.P., Kavlock, R.J., Tice, R.R., 2021. Chem. Res. Toxicol. 34, 189–216, https://doi.org/10.1021/acs.chemrestox.0c00264.

[46] Rotroff, D.M., Dix, D.J., Houck, K.A., Knudsen, T.B., Martin, M.T., McLaurin, K.W., Reif, D.M., Crofton, K.M., Singh, A.V., Xia, M., Huang, R., Judson, R.S., 2013. Environ. Health Perspect. 121, 7–14, https://doi.org/10.1289/ehp.1205065.

[47] Kar, S., Sepúlveda, M.S., Roy, K., Leszczynski, J., 2017. Chemosphere 184, 514–523, https://doi.org/10.1016/j.chemosphere.2017.06.024.

[48] Papa, E., Kovarich, S., Gramatica, P., 2013. SAR QSAR Environ. Res. 24, 333–349, https://doi.org/10.1080/1062936X.2013.773374.

[49] Papa, E., Kovarich, S., Gramatica, P., 2010. Chem. Res. Toxicol. 23, 946–954, https://doi.org/10.1021/tx1000392.

[50] Yang, X., Ou, W., Zhao, S., Xi, Y., Wang, L., Liu, H., 2021. ACS Sustain. Chem. Eng. 9, 5661–5672, https://doi.org/10.1021/acssuschemeng.1c00680.

[51] Natesan, S., Wang, T., Lukacova, V., Bartus, V., Khandelwal, A., Balaz, S., 2011. J. Chem. Inf. Model. 51, 1132–1150, https://doi.org/10.1021/ci200055s.

[52] Yang, W., Shen, S., Mu, L., Yu, H., 2011. Environ. Toxicol. Chem. 30, 2431–2439, https://doi.org/10.1002/etc.645.

[53] Organisation for Economic Co-operation and Development (OECD), 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en.

[54] Street, M.E., Audouze, K., Legler, J., Sone, H., Palanza, P., 2021. Int. J. Mol. Sci. 22, 933, https://doi.org/10.3390/ijms22020933.

[55] Cao, J., Lin, Y., Guo, L.-H., Zhang, A.-Q., Wei, Y., Yang, Y., 2010. Toxicology 277, 20–28, https://doi.org/10.1016/j.tox.2010.08.012.

[56] Chauhan, K.R., Kodavanti, P.R.S., McKinney, J.D., 2000. Toxicol. Appl. Pharmacol. 162, 10–21, https://doi.org/10.1006/taap.1999.8826.

[57] Cheek, A.O., Kow, K., Chen, J., McLachlan, J.A., 1999. Environ. Health Perspect. 107, 273–278, https://doi.org/10.1289/ehp.99107273.

[58] den Besten, C., Vet, J.J.R.M., Besselink, H.T., Kiel, G.S., van Berkel, B.J.M., Beems, R., van Bladeren, P.J., 1991. Toxicol. Appl. Pharmacol. 111, 69–81, https://doi.org/10.1016/0041-008X(91)90135-2.

[59] Gales, L., Almeida, M.R., Arsequell, G., Valencia, G., Saraiva, M.J., Damas, A.M., 2008. Biochim. Biophis. Acta. 1784, 512–517, https://doi.org/10.1016/j.bbapap.2007.11.014.

[60] Grimm, F.A., Lehmler, H.-J., He, X., Robertson, L.W., Duffel, M.W., 2013. Environ. Health Perspect. 121, 657–662, https://doi.org/10.1289/ehp.1206198.

[61] Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Visser, T.J., Van Velzen, M.J.M., Brouwer, A., Bergman, Å., 2008. Mol. Nutr. Food Res. 52, 284–298, https://doi.org/10.1002/mnfr.200700104.

[62] Hamers, T., Kortenkamp, A., Scholze, M., Molenaar, D., Cenijn, P.H., Weiss, J.M., 2020. Environ. Health Perspect. 128, 017015, https://doi.org/10.1289/EHP5911.

[63] Harju, M., Hamers, T., Kamstra, J.H., Sonneveld, E., Boon, J.P., Tysklind, M., Andersson, P.L., 2007. Environ. Toxicol. Chem. 26, 816–826, https://doi.org/10.1897/06-308R.1.

[64] Hill, K.L., Mortensen, Å.-K., Teclechiel, D., Willmore, W.G., Sylte, I., Jenssen, B.M., Letcher, R.J., 2018. Environ. Sci. Technol. 52, 1533–1541, https://doi.org/10.1021/acs.est.7b04617.

[65] Huang, K., Wang, X., Zhang, H., Zeng, L., Zhang, X., Wang, B., Zhou, Y., Jing, T., 2020. Environ. Sci. Technol. 54, 5437–5445, https://doi.org/10.1021/acs.est.9b05761.

[66] Legler, J., Cenijn, P.H., Malmberg, T., Bergman, A., Brouwer, A., 2002. Organohalogen Compd. 53–56, https://hdl.handle.net/1871.1/77c55c42-b5ae-4874-a14e-363929dcb267.

[67] Maia, F., Almeida, M. do R., Gales, L., Kijjoa, A., Pinto, M.M.M., Saraiva, M.J., Damas, A.M., 2005. Biochem. Pharmacol. 70, 1861–1869, https://doi.org/10.1016/j.bcp.2005.09.012.

[68] Marchesini, G.R., Meimaridou, A., Haasnoot, W., Meulenberg, E., Albertus, F., Mizuguchi, M., Takeuchi, M., Irth, H., Murk, A.J., 2008. Toxicol. Appl. Pharmacol. 232, 150–160, https://doi.org/10.1016/j.taap.2008.06.014.

[69] Marchesini, G.R., Meulenberg, E., Haasnoot, W., Mizuguchi, M., Irth, H., 2006. Anal. Chem. 78, 1107–1114, https://doi.org/10.1021/ac051399i.

[70] Meerts, I.A.T.M., van Zanden, J.J., Luijks, E.A.C., van Leeuwen-Bol, I., Marsh, G., Jakobsson, E., Bergman, Å., Brouwer, A., 2000. Toxicol. Sci. 56, 95–104, https://doi.org/10.1093/toxsci/56.1.95.

[71] Montaño, M., Cocco, E., Guignard, C., Marsh, G., Hoffmann, L., Bergman, Å., Gutleb, A.C., Murk, A.J., 2012. Toxicol. Sci. 130, 94–105, https://doi.org/10.1093/toxsci/kfs228.

[72] Ouyang, X., Froment, J., Leonards, P.E.G., Christensen, G., Tollefsen, K.-E., de Boer, J., Thomas, K.V., Lamoree, M.H., 2017. Chemosphere 171, 722–728, https://doi.org/10.1016/j.chemosphere.2016.12.119.

[73] Qin, W.-P., Li, C.-H., Guo, L.-H., Ren, X.-M., Zhang, J.-Q., 2019. Environ. Sci. Process. Impacts 21, 950–956, https://doi.org/10.1039/C9EM00095J.

[74] Ren, X.-M., Qin, W.-P., Cao, L.-Y., Zhang, J., Yang, Y., Wan, B., Guo, L.-H., 2016. Toxicol. 366–367, 32–42, https://doi.org/10.1016/j.tox.2016.08.011.

[75] Ren, X.-M., Yao, L., Xue, Q., Shi, J., Zhang, Q., Wang, P., Fu, J., Zhang, A., Qu, G., Jiang, G., 2020. Environ. Health Perspect. 128, 107008, https://doi.org/10.1289/EHP6498.

[76] Rosenmai, A.K., Winge, S.B., Möller, M., Lundqvist, J., Wedebye, E.B., Nikolov, N.G., Lilith Johansson, H.K., Vinggaard, A.M., 2021. Chemosphere 263, 127703, https://doi.org/10.1016/j.chemosphere.2020.127703.

[77] Sandau, C.D., Meerts, I.A.T.M., Letcher, R.J., McAlees, A.J., Chittim, B., Brouwer, A., Norstrom, R.J., 2000. Environ. Sci. Technol. 34, 3871–3877, https://doi.org/10.1021/es001134f.

[78] Simon, E., Bytingsvik, J., Jonker, W., Leonards, P.E.G., de Boer, J., Jenssen, B.M., Lie, E., Aars, J., Hamers, T., Lamoree, M.H., 2011. Environ. Sci. Technol. 45, 7936–7944, https://doi.org/10.1021/es2016389.

[79] Simon, E., van Velzen, M., Brandsma, S.H., Lie, E., Løken, K., de Boer, J., Bytingsvik, J., Jenssen, B.M., Aars, J., Hamers, T., Lamoree, M.H., 2013. Environ. Sci. Technol. 47, 8902–8912, https://doi.org/10.1021/es401696u.

[80] van den Berg, K.J., 1990. Chem. Biol. Interact. 76, 63–75, https://doi.org/10.1016/0009-2797(90)90034-K.

[81] Viluksela, M., Heikkinen, P., Ven, L.T.M. van der, Rendel, F., Roos, R., Esteban, J., Korkalainen, M., Lensu, S., Miettinen, H.M., Savolainen, K., Sankari, S., Lilienthal, H., Adamsson, A., Toppari, J., Herlin, M., Finnilä, M., Tuukkanen, J., Leslie, H.A., Hamers, T., Hamscher, G., Al-Anati, L., Stenius, U., Dervola, K.-S., Bogen, I.-L., Fonnum, F., Andersson, P.L., Schrenk, D., Halldin, K., Håkansson, H., 2014. PLoS One 9, e104639, https://doi.org/10.1371/journal.pone.0104639.

[82] Weiss, J.M., Andersson, P.L., Zhang, J., Simon, E., Leonards, P.E.G., Hamers, T., Lamoree, M.H., 2015. Anal. Bioanal. Chem. 407, 5625–5634, https://doi.org/10.1007/s00216-015-8736-9.

[83] Xi, Y., Yang, X., Zhang, H., Liu, H., Watson, P., Yang, F., 2020. Chemosphere 242, 125135, https://doi.org/10.1016/j.chemosphere.2019.125135.

[84] Yang, X., Ou, W., Xi, Y., Chen, J., Liu, H., 2019. Environ. Sci. Technol. 53, 7019–7028, https://doi.org/10.1021/acs.est.9b00218.

[85] Zhang, J., Begum, A., Brännström, K., Grundström, C., Iakovleva, I., Olofsson, A., Sauer-Eriksson, A.E., Andersson, P.L., 2016. Environ. Sci. Technol. 50, 11984–11993, https://doi.org/10.1021/acs.est.6b02771.

[86] Zhang, J., Kamstra, J.H., Ghorbanzadeh, M., Weiss, J.M., Hamers, T., Andersson, P.L., 2015. Environ. Sci. Technol. 49, 10099–10107, https://doi.org/10.1021/acs.est.5b01742.

[87] Kovarich, S., Papa, E., Li, J., Gramatica, P., 2012. SAR QSAR Environ. Res. 23, 207–220, https://doi.org/10.1080/1062936X.2012.657235.

[88] Good, P.I., Hardin, J.W., 2009. Common Errors in Statistics (And How to Avoid Them). Third edition, Wiley. https://doi.org/10.1002/9781118360125.

[89] Somack, R., Andrea, T.A., Jorgensen, E.C., 1982. Biochemistry 21, 163–170, https://doi.org/10.1021/bi00530a028.

[90] Nilsson, S., Rask, L., Peterson, P., 1975. J. Biol. Chem. 250, 8554–8563, https://doi.org/10.1016/S0021-9258(19)40795-3.

[91] Smith, D.S., 1977. FEBS Lett. 77, 25–27, https://doi.org/10.1016/0014-5793(77)80185-3.

[92] Wold, S., Esbensen, K., Geladi, P., 1987. Chemometr. Intell. Lab. 2, 37–52, https://doi.org/10.1016/0169-7439(87)80084-9.

[93] O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. J Cheminform 3, 33, https://doi.org/10.1186/1758-2946-3-33.

[94] Yap, C.W., 2011. J. Comput. Chem. 32, 1466–1474, https://doi.org/10.1002/jcc.21707.
[95] Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. J. Comput. Chem. 34, 2121–2132, https://doi.org/10.1002/jcc.23361.
[96] Todeschini, R., Consonni, V., Maiocchi, A., 1999. Chemometr. Intell. Lab. 46, 13–29, https://doi.org/10.1016/S0169-7439(98)00124-5.
[97] Tropsha, A., Gramatica, P., Gombar, V.K., 2003. QSAR Comb. Sci. 22, 69–77 https://doi.org/10.1002/qsar.200390007.
[98] Kier, L.B., Hall, L.H., 1990. Pharm. Res. 7, 801–807, https://doi.org/10.1023/A:1015952613760.
[99] Todeschini R, Consonni V, 2000. Handbook of molecular descriptors. Wiley. https://doi.org/10.1002/9783527613106.
[100] Moreau, G., Broto, P., 1980. Nouv. J. Chim. 4, 359–360.
[101] Hall, L.H., Kier, L.B., 1995. J. Chem. Inf. Comput. Sci. 35, 1039–1045, https://doi.org/10.1021/ci00028a014.

CHAPTER 3

In Silico Screening of Pharmaceuticals and Personal Care Products

The research and findings presented in this chapter draw heavily upon the data and methodologies first published in a co-authored, peer-reviewed article, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

The article is entitled *Global Assessment of Emerging Contaminant Removal in Wastewater Treatment Plants: In Silico Hazard Screening and Risk Evaluation* (see paper IV in list of publications) and it was published by Arianna Sgariboldi et al. in Toxics, Volume 13, 6, 2025 (https://doi.org/10.3390/toxics13010006).

The content herein constitutes an original synthesis and critical re-elaboration of that work, specifically adapted for the comprehensive scope of this dissertation.

3.1 Introduction

In the study entitled *Global Assessment of Emerging Contaminant Removal in Wastewater Treatment Plants: In Silico Hazard Screening and Risk Evaluation* [1] (paper IV in list of publications), a battery of twenty-four quantitative structure-activity relationship (QSAR) models, including those presented in Chapter 2 of this thesis, was applied to predict multiple properties and activities by a set of pharmaceuticals and personal care products (PPCPs).

PPCPs are recognised as a class of emerging contaminants given their continuous emission into the environment (mainly from human and veterinary consumption), their consequent widespread distribution, and the growing evidence of the adverse effects they can lead to humans and wildlife populations [2]. After use, PPCPs and their metabolites are excreted and enter the wastewater stream. A key pathway for their release into the environment is through wastewater treatment plants (WWTPs) effluents. Since many PPCPs are not efficiently removed by conventional wastewater treatment processes, WWTPs are a major source of contamination for aquatic bodies, underscoring the limitations of current water treatment technologies [3]. The ubiquitous presence of PPCPs in the environment is linked to a range of documented acute or chronic adverse effects on non-target organisms in the aquatic environment, including endocrine disrupting (ED)-related outcomes, physiological and behavioural changes, cellular and organ damage on the liver, kidney, and gills of fish [4]. A defining characteristic of PPCPs is that they are biologically active by design, since they have been synthetised to elicit specific physiological responses. As a consequence of this intrinsic bioactivity, effects by PPCPs can be observed even at low concentrations, often in the range of nanograms to micrograms per litre (ng/L to μg/L), at which they are found in the environment [4].

Beyond these specific biological impacts, many PPCPs exhibit properties like persistence (P), bioaccumulation (B), and mobility (M) in water [5,6]. In Europe, substances exhibiting these behaviours are classified as PBT or PMT. These properties are of major regulatory concern, to such extent that the CLP regulation was amended in 2023 by introducing new ad hoc hazard classes [7]. Under REACH regulation, PBT substances are currently identified as substances of very high concern (SVHC) [8].

First, this study provided an in-depth analysis of the removal efficiency (RE) of over 200 PPCPs in WWTPs situated in five continents, allowing for a comparison of the

performance of different treatment technologies. Then, a battery of QSAR models was applied to predict the potential PBT, PMT, and ED properties of PPCPs, in order to identify chemicals of greatest concern. Regarding ED properties, models for predicting binding to androgen receptor, to estrogen receptor, and T4-hTTR competing potency (i.e. those presented in Chapter 2 of this thesis) were applied. Finally, measured environmental concentrations and RE data were combined with the QSARs predictions to identify chemicals that may pose a risk to the aquatic environment.

3.2 Materials and methods

An extensive bibliographic search using Web of Science and Google Scholar databases was conducted to identify relevant papers published between 2014 and 2019. This search yielded 73 publications, from which 32 were selected based on strict criteria including detailed information on WWTPs and quantifiable data on chemical concentrations in WWTPs influents and effluents. From these publications, a total of 2034 individual data points for 251 different PPCPs were compiled. To ensure data reliability, a quality index (QI) was created and assigned to each data point based on three criteria: the clarity of chemical identification, the quality of RE identification, and the completeness of information about the WWTPs. Data points were scored and assigned to one of four quality classes: low, moderate, good, and excellent. A total of 282 data points (13.9% of the total) were categorised to have a QI "low" and were discarded. This filtering process excluded six PPCPs entirely, resulting in a final dataset of 245 unique PPCPs for subsequent analysis. The chemical information (i.e., names and CASRN numbers) was standardised, and molecular structures of PPCPs were represented as SMILES notations for application of the models.

A total of 76 WWTPs from the selected studies were analysed and categorised in WWTP1 and WWTP2 based on their level of technology. WWTP1 were plants using conventional primary (mechanical) and secondary (biological) treatments. WWTP2 were plants employing more advanced tertiary treatments, such as chemical-physical processes, advanced oxidation, or disinfection. In cases of ambiguity, a conservative worst-case approach was adopted, classifying plants with only advanced treatment method as WWTP1.

RE was either taken from the literature or calculated using a standard formula based on influent and effluent concentrations of PPCPs, preferentially using median values to avoid influence by outliers. A more complex formula accounting for chemical sorption to sludge was used when sufficient data was available. The final RE data for each compound was categorised into six classes and a consensus approach was applied to assign a single, representative RE class for WWTP1 and WWTP2.

SMILES notations of the 245 PPCPs were used as input for several freely available and regulatory-relevant software for the application of the QSARs, i.e. EPI SUITE v.4.1 [9], OPERA v.2.9 [10], the OECD QSAR Toolbox v.4.4 [11], and the QSAR-ME Profiler v.1.02 (available at https://dunant.dista.uninsubria.it/qsar/). The screening was

conducted on two levels, i.e. level 1 and level 2. In level 1, regulatory-preferred models and endpoints were used to assess persistence (e.g., BIOWIN to predict biodegradability), bioaccumulation (e.g., KOWWIN to predict log Kow), mobility (e.g., KOCWIN to predict log Koc), and toxicity (e.g., models in the QSAR-ME Profiler software to predict log LC50/EC50 for each aquatic trophic level, i.e. fish, daphnia, and algae) against established thresholds [12–14]. In level 2, supplementary evidence was provided by applying additional models implemented in the QSAR-ME Profiler v.1.02 and OPERA v.2.9. The potential risk (PR) to the aquatic environment by each PPCP was quantified by calculating the ratio between measured concentration of the PPCP in WWTP effluents and its predicted no-effect concentration (PNEC). PNEC values were calculated following official technical guidance, hence the lowest predicted acute toxicity value across the three trophic levels was divided by an assessment factor equal to 1000 [15]. A PPCP was flagged as posing a potential risk if any of its calculated PR values exceeded 1.

A particularly significant element of this study was the dedicated *in silico* screening for ED potential by PPCPs. A targeted set of validated QSAR models was applied to predict the interaction of the 245 PPCPs with key receptors and proteins covering three major endocrine modalities (i.e., estrogen, androgen, and thyroid). The CERAPP (collaborative estrogen receptor activity prediction project) consensus model [16], which is available in the OPERA software, was applied to predict the ability of PPCPs to bind to the estrogen receptor (ER). The COMPARA (collaborative modelling project for androgen receptor activity) consensus model [17], also available in the OPERA software, was applied to predict the ability of PPCPs to bind to the androgen receptor (AR). Finally, the models presented in Chapter 2 of this thesis [18] and available in the QSAR-ME Profiler software were applied to assess the potential TH system disruption by PPCPs.

3.3 Results and discussion

The analysis confirmed that advanced WWTPs (i.e., WWTP2 in this study) were significantly more effective at removing PPCPs than conventional WWTPs (i.e., WWTP1 in this study). Indeed, about 60% of PPCPs achieved moderate to excellent removal in WWTP2, compared to only about 40% in WWTP1. However, certain chemical classes, notably antibiotics and psychiatric drugs, proved to be highly resistant to treatment even in advanced plants. Furthermore, many PPCPs that were effectively removed by WWTP2 were found in regions like Africa and parts of Asia, where WWTP1 are more common. This highlighted a disparity in wastewater treatment infrastructure and pointed out the need for technology investments in developing countries to mitigate the release of hazardous chemicals into the environment.

Application of the QSARs presented in Chapter 2 led to the results reported in Table 3.1. As shown, most of the screened PPCPs fell inside the applicability domains (ADs) of the QSARs, although with the RLBA-based QSAR being less capable to provide reliable predictions.

Table 3.1. Summary of the screening results from the application of the QSARs presented in Chapter 2.

Model	Inside AD	Outside AD
ANSA	234	11
FITC-T4	211	34
RLBA	148	97

A total of 125 out of 245 PPCPs were reliably predicted as potential T4-hTTR competitors by all of the three QSAR models. Figure 3.1 shows the distribution of these compounds across different classes of PPCPs. The most represented were antibiotics, followed by drugs and metabolites. The category "Others" includes nineteen different classes of PPCPs. Among these, the most represented were antiepileptics and metabolites, hormones and steroids, lipid regulators, parabens, and sedatives-hypnotics-anxiolytics, including four compounds in each class.

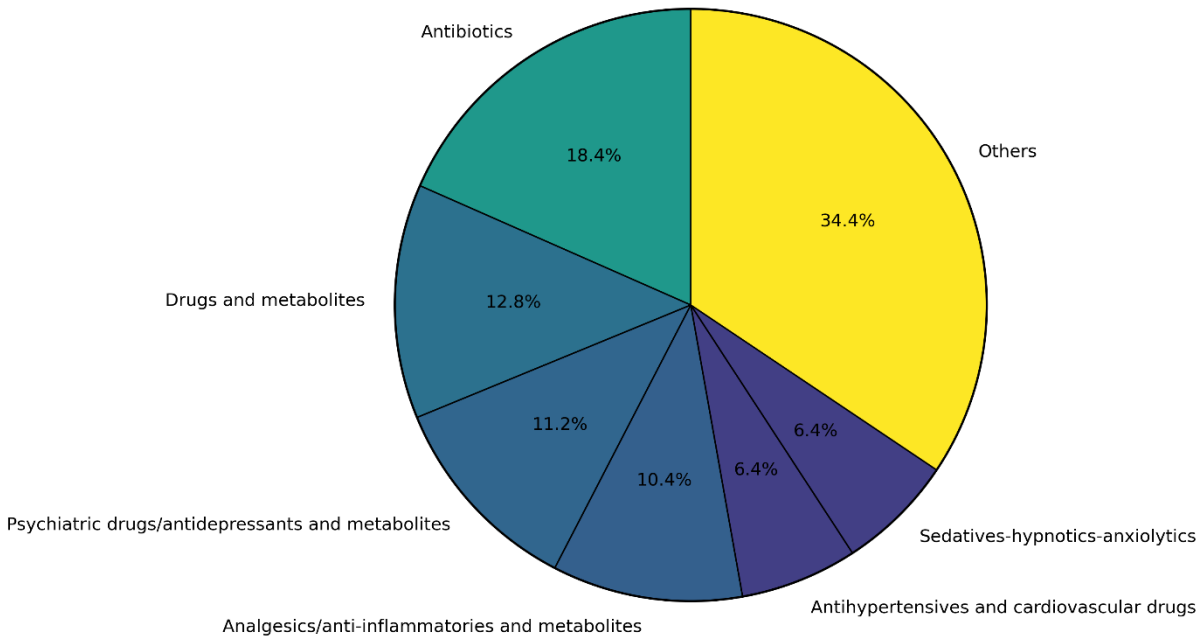

Figure 3.1. Percentage of chemicals predicted inside the AD of all the three QSARs presented in Chapter 2, grouped by class.

The *in silico* screening identified 16 PPCPs as potentially PBT or PMT. They included four antibiotics, six psychiatric drugs, and two opioids. These PPCPs pose a threat to environmental and human health, and should be prioritised for further investigation and scrutiny. The most compelling finding from this screening was that all 16 of the PPCPs identified as potentially PBT or PMT were also flagged as potential endocrine disruptors. For example: citalopram, sertraline, triclocarban, and miconazole were all flagged as potential AR binders and strong hTTR binders. Regarding T4-hTTR competing potency predictions, eight PPCPs were identified as strong hTTR binders, seven as moderate hTTR binders, and one as weak hTTR binder.

Analysis of the risk identified 77 PPCPs linked to 264 potential risk cases (i.e., PR > 1). A striking and important finding was that while over 60% of these cases were associated with chemicals that are poorly removed, approximately 40% of the risk cases were for chemicals considered well-removed, i.e., with high or excellent RE. This demonstrated that effective RE alone does not guarantee safety, and the residual concentration of a substance can still be sufficient to pose a risk.

3.4 Conclusions

In this study, a robust and integrated methodology for screening and prioritising emerging contaminants like PPCPs was conducted. By combining RE data with *in silico* hazard predictions, relevant insights for scientists and regulators were provided. In particular, specific classes of PPCPs, such as antibiotics and psychiatric drugs, are highly resistant to wastewater treatment and are therefore continuously released and present in the environment. Based on screening results, a total of 16 PPCPs were identified as being of high concern due to their potential PBT, PMT, and ED properties, and should undergo further investigation. Environmental risk is not solely dictated by RE, since even PPCPs that are efficiently removed by WWTPs can pose a threat if they are sufficiently toxic, highlighting the need to reduce their release at the source. In addition, based on the application of QSARs to identify potential TH system disruptors, multiple classes were identified as particularly dangerous, with antibiotics being the most concerning one. This study is also affected by the following known limitation. The QSAR models used in this work were developed for neutral organic compounds, therefore they are not able to take into account the ionization state of chemicals. Given that, the models were applied on neutralised structures of the PPCPs. However, most PPCPs are ionizable organic chemicals, meaning that their properties are pH-dependent. The application of QSARs on neutralised structures may not fully capture the environmental behaviour and/or toxicity of PPCPs.

3.5 REFERENCES

[1] Sgariboldi, A., Posté, E., Chirico, N., Sangion, A., Evangelista, M., Morosini, C., Re, A., Torretta, V., Papa, E., 2025. Toxics 13, 6, https://doi.org/10.3390/toxics13010006.

[2] Osuoha, J.O., Anyanwu, B.O., Ejileugha, C., 2023. J. Hazard. Mater. Adv. 9, 100206, https://doi.org/10.1016/j.hazadv.2022.100206.

[3] Kumar, M., Sridharan, S., Sawarkar, A.D., Shakeel, A., Anerao, P., Mannina, G., Sharma, P., Pandey, A., 2023. Sci. Total Environ. 859, 160031, https://doi.org/10.1016/j.scitotenv.2022.160031.

[4] Rehman, M.U., Nisar, B., Mohd Yatoo, A., Sehar, N., Tomar, R., Tariq, L., Ali, S., Ali, A., Mudasir Rashid, S., Bilal Ahmad, S., Aldossari, R.M., 2024. Sep. Purif. Technol. 342, 126921, https://doi.org/10.1016/j.seppur.2024.126921.

[5] Huang, C., Jin, B., Han, M., Yu, Y., Zhang, G., Arp, H.P.H., 2021. J. Hazard. Mater. Lett. 2, 100026, https://doi.org/10.1016/j.hazl.2021.100026.

[6] Zenker, A., Cicero, M.R., Prestinaci, F., Bottoni, P., Carere, M., 2014. Bioaccumulation and biomagnification potential of pharmaceuticals with a focus to the aquatic environment. J. Environ. Manage. 133, 378–387, https://doi.org/10.1016/j.jenvman.2013.12.017.

[7] European Commission, 2022. Commission Delegated Regulation (EU) 2023/707 of 19 December 2022 amending Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and packaging of substances and mixtures (Text with EEA relevance).

[8] REACH Regulation (EC) No. 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals. Brussels, Belgium., 2006.

[9] US EPA, 2015. Estimation Programs Interface (EPI) Suite.

[10] Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. J. Cheminform. 10, 10, https://doi.org/10.1186/s13321-018-0263-1.

[11] OECD QSAR Toolbox version 4.4. https://qsartoolbox.org/ (last accessed on 06 October 2025).

[12] European Chemicals Agency (ECHA), 2017. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB Assessment. https://doi.org/10.2823/312974.

[13] Neumann, M., Schliebner, I., 2019. Umweltbundesamt (UBA) TEXTE 127/2019 Protecting the Sources of Our Drinking Water: The Criteria for Identifying Persistent, Mobile and Toxic (PMT) Substances and Very Persistent and Very Mobile (vPvM) Substances Under EU Regulation REACH (EC) No 1907/2006. https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-11-29_texte_127-2019_protecting-sources-drinking-water-pmt.pdf (last accessed on 06 October 2025).

[14] Arp, H.P.H., Hale, S.E., Schliebner, I., Neumann, M., 2023. Umweltbundesamt (UBA) TEXTE 21/2023 Prioritised PMT/vPvM Substances in the REACH Registration Database. https://www.umweltbundesamt.de/publikationen/prioritised-pmtvpvm-substances-in-the-reach (last accessed on 06 October 2025).

[15] De Bruijn, J., Hansen, B., Johansson, S., Luotamo, M., Munn, S., Musset, C., Olsen, S., Olsson, H., Paya-Perez, A., Pedersen, F., Rasmussen, K., Sokull-Kluttgen, B., 2003. Technical Guidance Document on Risk Assessment: Part II. https://publications.jrc.ec.europa.eu/repository/handle/JRC23785 (last accessed on 06 October 2025).

[16] Mansouri, K., Abdelaziz, A., Rybacka, A., Roncaglioni, A., Tropsha, A., Varnek, A., Zakharov, A., Worth, A., Richard, A.M., Grulke, C.M., Trisciuzzi, D., Fourches, D., Horvath, D., Benfenati, E., Muratov, E., Wedebye, E.B., Grisoni, F., Mangiatordi, G.F., Incisivo, G.M., Hong, H., Ng, H.W., Tetko, I.V., Balabin, I., Kancherla, J., Shen, J., Burton, J., Nicklaus, M., Cassotti, M., Nikolov, N.G., Nicolotti, O., Andersson, P.L., Zang, Q., Politi, R., Beger, R.D., Todeschini, R., Huang, R., Farag, S., Rosenberg, S.A., Slavov, S., Hu, X., Judson, R.S., 2016. Environ. Health Perspect. 124, 1023–1033, https://doi.org/10.1289/ehp.1510267.

[17] Mansouri, K., Kleinstreuer, N., Abdelaziz, A.M., Alberga, D., Alves, V.M., Andersson, P.L., Andrade, C.H., Bai, F., Balabin, I., Ballabio, D., Benfenati, E., Bhhatarai, B., Boyer, S., Chen, J., Consonni, V., Farag, S., Fourches, D., García-Sosa, A.T., Gramatica, P., Grisoni, F., Grulke, C.M., Hong, H., Horvath, D., Hu, X., Huang, R., Jeliazkova, N., Li, J., Li, X., Liu, H., Manganelli, S., Mangiatordi, G.F., Maran, U., Marcou, G., Martin, T., Muratov, E., Nguyen, D.-T., Nicolotti, O., Nikolov, N.G., Norinder, U., Papa, E., Petitjean, M., Piir, G., Pogodin, P., Poroikov, V., Qiao, X., Richard, A.M., Roncaglioni, A., Ruiz, P., Rupakheti, C., Sakkiah, S., Sangion, A., Schramm, K.-W., Selvaraj, C., Shah, I., Sild, S., Sun, L., Taboureau, O., Tang, Y., Tetko, I.V., Todeschini, R., Tong, W., Trisciuzzi, D., Tropsha, A., Van Den Driessche, G., Varnek, A., Wang, Z., Wedebye, E.B., Williams, A.J., Xie, H., Zakharov, A.V., Zheng, Z., Judson, R.S., 2020. Environ. Health Perspect. 128, 027002, https://doi.org/10.1289/EHP5580.

[18] Evangelista, M., Chirico, N., Papa, E., 2024. In silico models for the screening of human transthyretin disruptors. J. Hazard Mater. 480, 136188, https://doi.org/10.1016/j.jhazmat.2024.136188.

CHAPTER 4

New QSAR Models to Predict Human Transthyretin Disruption by Per- and Polyfluoroalkyl Substances (PFAS): Development and Application

Marco Evangelista, Nicola Chirico, and Ester Papa

Published in Toxics, Volume 13, 590, 2025
https://doi.org/10.3390/toxics13070590

Abstract

Per- and polyfluoroalkyl substances (PFAS) are of concern because of their potential thyroid hormone system disruption by binding to human transthyretin (hTTR). However, the amount of experimental data is scarce. In this work, new classification and regression quantitative structure-activity relationship (QSAR) models were developed to predict the hTTR disruption based on experimental data measured for 134 PFAS. Bootstrapping, randomization procedures, and external validation were used to check for overfitting, to avoid random correlations, and to evaluate the predictivity of the QSARs, respectively. The best QSARs were characterized by good performances (e.g., training and test accuracies in classification of 0.89 and 0.85, respectively; R^2, Q^2_{loo}, and Q^2_{F3} in regression of 0.81, 0.77, and 0.82, respectively) and significantly broader domains compared to the few existing similar models. The application of QSARs to the Organisation for Economic Co-operation and Development (OECD) List of PFAS allowed for the identification of structural categories of major concern, such as per- and polyfluoroalkyl ether-based, perfluoroalkyl carbonyl, and perfluoroalkane sulfonyl compounds. Forty-nine PFAS showed a stronger binding affinity to hTTR than the natural ligand T4. Uncertainty quantification for each model and prediction further enhanced the reliability assessment of predictions. The implementation of the new QSARs in non-commercial software facilitates their application to support future research efforts and regulatory actions.

4.1 Introduction

Per- and polyfluoroalkyl substances (PFAS) are a large and largely heterogeneous class of human-made compounds, whose strong carbon–fluorine bonds in their structure provide them unique properties (e.g., amphipathic nature, chemical and thermal stability) that have led to their widespread use in different industrial and consumer applications [1]. However, many PFAS and their terminal transformation products are characterized by high persistence in environmental matrices (e.g., surface and groundwater, soils, sediments, atmosphere) due to their resistance to biotic and abiotic degradation processes under environmental conditions [1,2]. Furthermore, some PFAS have been reported to be mobile in the environment, and to bioaccumulate in living organisms [2]. The combination of persistence and mobility of PFAS results in their global contamination [2], leading to increasing exposure of humans and wildlife through multiple pathways, including oral ingestion of contaminated water and food, inhalation of airborne particles, dermal contact with environmental matrices, and consumer and personal care products [3]. Exposure to PFAS has been linked to several serious diseases in mammals, reptiles, fish, birds, and amphibians, including neurotoxicity, immunotoxicity, reproductive impairment, and endocrine disruption [4,5]. The lowest observed effect concentrations (LOECs) of different PFAS have been reported mainly in the ng/mL—μg/mL range in biological matrices, depending on the species and the adverse effect under investigation [5]. PFAS have also been shown to cause toxic effects on terrestrial and aquatic invertebrates, with lethal and effect concentrations to 50% of the population (LC50 and EC50, respectively) mainly reported in the mg/kg or mg/L range [6]. The ubiquitous presence of PFAS and the serious threats they pose raise concerns for human health and the environment that need to be addressed.

Endocrine disruption (ED) may occur through the interference of both legacy and emerging PFAS with multiple molecular targets encompassed in the hypothalamic–pituitary–thyroid (HPT) axis [7]. In mammalian species, the HPT axis is responsible for regulating thyroid hormones (THs) homeostasis [8], the proper functioning of which is critical as THs play a key role in multiple biological functions during both fetal and post-natal life stages [9–12]. One biological mechanism by which xenobiotics (e.g., PFAS) can interfere with the physiological functions of the HPT axis is the competition with the TH thyroxine (T4) for binding to the human TH distributor protein transthyretin (hTTR), which has been identified as a critical molecular initiating event (MIE) in adverse outcome pathway (AOP) networks for TH system

disruption [13,14]. hTTR is involved in a variety of biological functions, including the regulation of abnormal changes in the serum levels of free THs, and the mediation of T4 delivery from blood to cerebrospinal fluid across critical barriers, such as the blood–brain barrier and the placenta, during fetal development [15]. PFAS exposure of vulnerable populations, such as pregnant women, is thus a critical issue [16] as THs regulate brain differentiation and central nervous system formation [17], and play a key role in the metabolism, differentiation, and development of the placenta [18]. Since the embryo/fetus relies entirely on maternal THs during the early stages of gestation, any disruption in THs supply can have significant, even irreversible, consequences that can extend beyond neonatal life [19]. Furthermore, a recent study advanced the hypothesis of the potential multi-transgenerational effects of PFAS on the thyroid axis [20]. For these reasons, the fast identification of substances exhibiting this type of toxicity, defined as thyroid hormone system-disrupting chemicals (THSDCs), is urgently needed [21].

Although associations between PFAS and TH system disruption have been proven, research studies still predominantly target known legacy PFAS. For many others, and in particular for short-chain and emerging variants, such information is heavily limited or absent [22–24]. This creates alarming gaps in the understanding of PFAS toxicity and generates substantial challenges in their evaluation on an individual basis [22–24]. In this context, the development and application of new approach methodologies (NAMs), including quantitative structure–activity relationship (QSAR) models, is being strongly promoted by authorities [25], intergovernmental organizations [26], and the scientific community [27–29] to accelerate the ED assessment of substances, including potential TH system disruption by PFAS [22,23,30,31], and to facilitate data gap filling, prioritization, and grouping strategies [32,33]. To our knowledge, only three studies have so far proposed QSAR models to specifically predict the potential hTTR disruption by PFAS [34–36]. However, these studies were affected by several limitations. The first was the use of commercial software, such as HyperChem (Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA) [37], Dragon (version 5.5 and version 6.0) [38], and alvaDesc [39], to optimize 3D molecular structures and/or calculate molecular descriptors, which may limit the application of these models. In addition, each study used small datasets to develop QSARs, resulting in models with a weak applicability domain (AD), thus limiting the reliability of predictions for a wider range of chemical structures and responses. In particular, the QSARs proposed by Kar et al. [34] and Kovarich et al. [35] were based on a dataset of experimental hTTR binding affinity data measured for 24 PFAS using the radiolabeled

[125I]-T4 *in vitro* binding assay (RLBA) [40], which is now not considered as suitable as competitive fluorescence displacement assays and is thus currently not being validated by the European Commission's European Union Reference Laboratory for alternatives to animal testing (EURL ECVAM) to measure the binding to hTTR [21]. Sosnowska et al. [36] proposed QSARs using a dataset of 44 PFAS for the prediction of the relative potency factor (RPF), calculated as the ratio of the potency of a specific PFAS to the toxic potency of perfluorooctanoic acid. However, the use of perfluorooctanoic acid as the reference compound does not actually reflect the ability of PFAS to compete with T4 for binding to hTTR. In addition, experimental data were measured using the TTR-TRβ CALUX assay. Analogously to the RLBA, this assay is currently not being validated by the EURL ECVAM [21]. Although these QSARs have provided valuable insights, their limitations related to the ADs coverage, the endpoints studied, the dimension of the datasets, the use of proprietary descriptors, and their availability for application emphasize the need for further work in this area. In addition, the availability of multiple models addressing different structural and response domains is strongly encouraged, allowing for the use of the consensus approach to improve the predictive ability of QSARs.

In this work, new classification and regression QSAR models are proposed and applied in a sequential approach with the aim of providing tools for the qualitative and quantitative screening of potential hTTR disruptors. The new models are intended to be applied to first identify hTTR-binding PFAS (by applying the classification QSAR) and then to quantify their T4-hTTR competing potency (by applying the regression QSAR). This work is an innovation compared to our previous regression models for the prediction of hTTR disruption [41], as it is specifically designed to address PFAS. This work introduces significant innovations, as described below, to overcome the limitations affecting the aforementioned QSARs [34–36]. The overarching aim is to address current gaps in the field by enhancing the transparency and robustness of QSAR models, and the reliability of QSAR predictions, in order to boost confidence in their use and promote their wider application, as well as to accelerate TH system disruption assessment of a class of priority substances like PFAS. To this end, a newly published dataset [42] was used in this study to generate the models and to provide external validation. The use of this dataset offers four main distinct advantages compared to those modelled in the foreign QSARs. First, it includes experimental hTTR binding affinities consistently measured for 134 heterogeneous PFAS. Its size is about three times larger than the largest considered in the previous studies [34–36]. Second, this dataset is sufficiently large to demonstrate

the robustness and predictive ability of the QSARs, and particularly to perform rigorous statistical procedures to detect and avoid overfitting and random correlations. It is worth highlighting that no comparable procedures were applied to ascertain that overfitting did not take place in the previous studies [34–36]. The size of a modelled dataset influences the statistical validation procedures that should be carried out, according to the Organisation for Economic Co-operation and Development (OECD) principles for QSAR development and validation [43]. Third, unlike existing models, this work uses data homogeneously measured with the 8-anilino-1-naphtalenesulfonic acid (ANSA)-based binding *in vitro* assay [42], which is a fluorescence-based competitive displacement assay that has been identified as a powerful method for identifying potential THSDCs [44] and is currently being validated by the EURL ECVAM to measure the hTTR binding of chemicals [21]. Fourth, the dataset was built by Degitz and colleagues by selecting PFAS by means of a category-based approach to ensure structural diversity [42]. This strategic selection maximized the chemical space coverage within the PFAS family for training predictive models. An additional significant innovation introduced in this study is the quantification of the uncertainty associated with each model and prediction. Generally, whether a QSAR can provide reliable or unreliable predictions is based on the AD defined on the information included in the training set [43]. However, multiple approaches are available to define the AD, which can vary in terms of the constraint degree impacting the reliability of predictions [43]. In this work, beyond the definition of an AD for each model, the uncertainty quantification is introduced to further enhance the reliability assessment of predictions to improve their confidence. Finally, the foreign QSARs relied on proprietary descriptors, which limits their broader application. In contrast, the models proposed in this work rely on descriptors calculated by non-commercial software [45]. Furthermore, the new QSARs have been implemented in the non-commercial software QSAR-ME Profiler beta version 1.02 (freely available for download at the authors' website https://dunant.dista.uninsubria.it/qsar/). This marks a clear advantage, as the QSARs are made freely available to scientists to aid the assessment of the hTTR disruption by PFAS from their molecular structure, providing not only a clear quantification of their ADs, but also uncertainty in predictions, which is not common in other QSARs. Finally, a case study is proposed to show how the sequential application of classification and regression QSARs can be used to screen large datasets of PFAS, such as the OECD List published in 2018 [46]. This list contains a comprehensive inventory of 4730 PFAS for which the potential hazards are still largely unknown [47].

4.2 Materials and methods

4.2.1 Modelled datasets and data curation

A dataset was retrieved from a newly published study by the United States Environmental Protection Agency (US EPA) [42], which included hTTR binding affinity values for 134 structurally heterogeneous PFAS. The data were measured using the ANSA-based binding *in vitro* assay, a fluorescence-based competitive displacement assay that has been identified as a powerful methodology for the identification of potential THSDCs [44]. The curation of data led to the exclusion of a total of 11 salts and organometals [48]. In order to develop the classification QSAR, the remaining 123 PFAS were classified as active if the median activity was greater than or equal to 50% (74 PFAS), and as weak/not active if the median activity was smaller than 50% (49 PFAS). Moreover, the active compounds were a priori defined as positive (Class A), while the weak/not active were defined as negative (Class I). Two distinct values of median activity were reported for two PFAS: 14.1% and 10.2% for 1H,1H-Perfluorooctylamine (Chemical Abstracts Service Registration Number, or CASRN, 307-29-9); 93.5% and 93.4% for 1,6-Diiodoperfluorohexane (CASRN 375-80-4). Nevertheless, the presence of these distinct values had no effect on the classification of these compounds. The development of the regression QSAR was based exclusively on PFAS with a quantitative hTTR binding affinity value (quantified in terms of half-maximal effect concentration), which resulted in 68 unique compounds. The modelled endpoint was the logarithm of the relative competitive potency (RP), which is defined as the ratio between the binding affinity of T4 and the binding affinity of a PFAS with hTTR. RP has been used in previous studies to quantify the ability of compounds to compete with T4 for binding to hTTR (i.e., T4-hTTR competing potency) [41,49–52]. For one compound (1,6-Diiodoperfluorohexane, CASRN 375-80-4), two distinct values of hTTR binding affinity were reported (1.712 μM and 1.848 μM); the arithmetic mean of the corresponding RP values was log-transformed and assigned to the compound of interest. The modelled datasets are reported in Supplementary Materials S1 (Tables S1 and S2).

4.2.2 Calculation of molecular descriptors and dataset splitting for external validation

The CASRN of each PFAS was used as input in the US EPA CompTox Chemicals Dashboard [53] to download simplified molecular input line entry system (SMILES) notations, which encode for the molecular structures. To ensure consistency, SMILES notations were canonicalized using Open Babel software v. 2.4.1 [54] and used as input in PaDEL-Descriptor software v. 2.21 [45] for the calculation of fingerprints, one-dimensional and two-dimensional theoretical molecular descriptors. Prior to modelling, an in-house R script (an algorithm previously published by our research group [55]) was used to filter the molecular descriptors in order to reduce useless and/or redundant information. Specifically, descriptors with low variance (i.e., constant value for more than 80% of the compounds), or exhibiting a pairwise correlation > 0.95, or with ranges larger than two orders of magnitude units, were excluded. In order to assess the predictive ability of the models on PFAS not used to train the models, each dataset was split into a training set for QSAR development and a test set for its external validation.

As the literature dataset [42] was built by selecting PFAS through a category-based approach to ensure broad structural diversity within the PFAS family, the splitting "by structure" procedure, already suggested in another study [56], was used to keep this structural diversity across both the training and test sets. This procedure first involved conducting a principal component analysis (PCA) [57] on the dataset using the molecular descriptors as input variables. Then, PFAS were ranked according to their scores along the first component. Based on this ranking, two-thirds of the PFAS were assigned to the training sets, and the remaining one-third was assigned to the test sets. Regarding the dataset used for classification, the splitting procedure was performed independently for each activity class. Finally, the molecular descriptors of the compounds included in the training sets were filtered to exclude redundant and useless information [55].

4.2.3 QSAR models development

4.2.3.1 Classification-based QSARs

Linear discriminant analysis (LDA) was used as the modelling algorithm. The variable subset selection was performed by applying the step-up procedure previously proposed by Rücker and colleagues [58]. This procedure was applied by means of an in-house developed R script tailored to perform this task [55]. The step-up procedure is described as a sort of stepwise selection [58], and it was chosen over other methods of variable selection (e.g., genetic algorithms) because it allows for the calculation of nested bootstrapped cross-validation in a smaller computational time. This method resulted in populations of the best LDA models ranked by their misclassification rate (MR), which is defined as the percentage of incorrect predictions (i.e., false positives (FPs) and false negatives (FNs) out of the total number of predictions). A linear scoring equation is provided for each class (i.e., Class A and Class I). A compound is assigned to the class associated with the equation that returns the higher score. The overfitting of the variable selection procedure [59] was checked by means of the leave-one-out bootstrap method [60] and evaluated in terms of bootstrapped MR (i.e., $MR_{BOOTSTRAP}$). The flattening or the increase of the $MR_{BOOTSTRAP}$, for an increasing number of modelling descriptors, is indicative of possible overfitting. The quality of the selected models and their predictive ability were evaluated using the following metrics: MR and accuracy (ACC), sensitivity (SN), specificity (SP), and precision (P). The analysis was further supported by the receiver operating characteristic (ROC) curve and area under the curve (AUC). Moreover, in order to minimize the possibility of developing models with coincidental relationships between the response and the descriptors, the probability of coincidental relationship was estimated by performing the step-up procedure for variable subset selection 100 times, using randomized descriptors [58] within their range of values, both considering the nature of the descriptors (i.e., discrete, continuous, binary) or not [55]. Additional details regarding the step-up algorithm and the formulae used to calculate the classification metrics listed above are provided in Supplementary Materials S2.

4.2.3.2 Regression-based QSARs

Multiple linear regression (MLR) by means of ordinary least squares (OLS) was used as the modelling algorithm. The variable subset selection was performed by selecting

the best combinations of modelling variables using the step-up procedure as in Section 2.3.1. In this case, the step-up procedure resulted in populations of the best MLR models ranked by their coefficient of determination (R^2). The overfitting of the variable selection procedure [59] was checked by means of the leave-one-out bootstrap, as outlined in Section 2.3.1, and evaluated in terms of bootstrapped mean absolute error (i.e., $MAE_{BOOTSTRAP}$). The probability of coincidental relationships between the molecular descriptors and the response was determined as outlined in Section 2.3.1. The evaluation of the models' fitting and internal robustness was conducted by using several metrics, i.e., R^2, MAE, and leave-one-out cross-validated R^2 (Q^2_{LOO}). The calculations of these metrics are reported in Supplementary Materials S2. Furthermore, the Y-scrambling procedure (50 iterations) was carried out to evaluate chance correlation between the descriptors and the response of the selected model, in terms of the average of R^2 (R^2_{YS}). Low R^2_{YS} values are observed in the absence of chance correlation among the model descriptors and the response. Finally, the plot of the residuals was generated to graphically verify the homoscedasticity of the residuals in prediction. As described in Section 2.3.1, the overall procedure was performed using an in-house developed script [55].

4.2.3.3 External validation

The predictive ability of the optimal classification and regression models was evaluated on the external test sets identified through the splitting procedure described in Section 4.2.2. It is important to highlight that the external test sets were not used for the training of the classification and the regression models. Rather, the external validation using the external test sets was performed after the selection of the best models on the basis of the assessment of their fitting and robustness quantified on the training set chemicals. The external predictivity of the selected models was quantified on the basis of predictions generated for the test set. Regarding the classification model, the external predictivity was quantified using ACC, SN, SP, and P, with the support of the ROC curve and AUC. Regarding the regression model, the external predictivity was quantified using the MAE and the external Q^2_{F3} [61].

4.2.4 Applicability domain

4.2.4.1 LDA-QSARs applicability domain

The AD of LDA-QSAR models was defined in terms of chemical structure and of post probabilities of the classification event. Specifically, a chemical structure was considered as an outlier if its distance, measured as the average of the three nearest cos α neighbors [62,63], was smaller than the 0.95 quantile of the k-nearest neighbors within the distribution of all training set distances [64]. The endpoint domain [62] was defined by post probability thresholds, which were selected on an arbitrary basis (i.e., for external predictions, the classification was considered uncertain if the post probability fell between 0.25 and 0.75). Following the application of the model, the reliability of each prediction was subjected to further evaluation by comparing its uncertainty, as estimated by Shannon entropy, with the maximum uncertainty calculated for the training set. Predictions with uncertainties within the maximum uncertainty calculated for the training set were considered reliable. Further details on the calculation of Shannon entropy values are provided in Supplementary Materials S2.

4.2.4.2 MLR-QSARs applicability domain

The AD of MLR-QSAR models was quantified using the leverage approach, with the graphical support of the Williams plot for the identification of structural and/or response outliers [43]. The leverage values (i.e., the hat matrix diagonal elements, defined in Supplementary Materials S2) of the compounds, which are a measure of their distance from the centroid of the model, were plotted on the x-axis of the Williams plot. The cut-off value h^* is defined as $3 \times (p + 1)/n$, where p is the number of model descriptors and n is the number of compounds included in the training set. The leverage value is indicative of the influence of a compound on the model and the reliability of its prediction. Compounds with leverage values exceeding h^* were considered to be structural outliers (predictions become less reliable as the leverage distance increases). The standardized residuals, plotted on the y-axis of the Williams plot, are a measure of the response AD. Compounds with a standardized residual exceeding ±2.5 standard deviation units were considered as response outliers. After the application of the models, the reliability of each prediction was subject to further assessment by comparing their uncertainty (i.e., prediction interval) with the maximum uncertainty calculated for the training set, as well as by comparing the predicted values with the experimental range of the response in the training set.

Reliable predictions have uncertainties within the maximum uncertainty calculated for the training set and predicted values within the experimental range of the response in the training set. Details on the calculation of prediction intervals are provided in Supplementary Materials S2.

4.2.5 OECD List of PFAS

The OECD List of PFAS, published in 2018 [46], was used in this work as the basis of a case study to demonstrate the application of the here-proposed LDA-QSAR and MLR-QSAR models to a substantial set of PFAS. This OECD List is an update of a previous data collection by the OECD, published in 2007 [65], to which seventeen publicly accessible information sources were added, and data curation was performed to include only substances with a defined CASRN [46]. Consequently, the OECD List used in this study [46] originally consisted of 4730 different PFAS (in terms of structures, applications, and regulatory status). Prior to the application of the QSARs, the OECD List underwent further data curation to remove polymers, mixtures, salts, organometals, and charged structures, as these chemicals are unsuitable for the application of the proposed models. Furthermore, compounds with ambiguous chemical identifiers, as well as those already included in the modelled datasets, were excluded. Stereoisomers were considered as duplicates in their non-chiral form, since QSAR models generated from simple bidimensional structures do not accurately reflect the spatial conformation of those compounds due to missing information on stereochemistry (for further information, see Table S3 in Supplementary Materials S1). The CASRNs of the remaining PFAS were used as input in the US EPA CompTox Chemicals Dashboard [53] to download the SMILES notation of their structures. The SMILES notations were subsequently canonicalized with Open Babel software v. 2.4.1 [54]. No SMILES notations were available for 62 compounds. The overall curation procedure led to a final dataset consisting of 2934 different neutral organic PFAS, including 53 distinct non-chiral forms of 109 stereoisomers that were initially included in the OECD List. The final dataset is reported in Supplementary Materials S1 (Table S4). The LDA-QSAR model was initially applied to classify PFAS in the curated OECD List as active or weak/not active. Compounds that fell outside the AD of the model, as described in Section 4.2.4.1, were removed. Subsequently, the MLR-QSAR model was applied to predict the T4-hTTR competing potencies (expressed as Log RP, as described in Section 4.2.1) of PFAS identified as active by the LDA-QSAR. As described in Section 4.2.4.2, compounds that fell outside the AD of the model were removed prior

to the analysis of the results. The value Log RP ≥ −1.26, suggested in the literature [52], was used as the threshold to identify strong hTTR binders among those PFAS screened as active by the LDA model.

4.3 Results and discussion

4.3.1 LDA-QSAR

PCA [57] was performed to study the structural space of the full dataset composed of 123 PFAS prior to modelling. Nine halogenated PFAS (where halogen atoms are intended to be bromine or iodine) were identified as structurally dissimilar to the remaining compounds. In a preliminary modelling attempt using the full dataset, two out of these nine compounds, i.e., heptafluorobutyl iodide (CASRN 374-98-1) and 1,6-dibromododecafluorohexane (CASRN 918-22-9), which is the only brominated PFAS in the dataset, were repeatedly misclassified. It was verified that these outliers had a significant impact on the performances of the models, and they were consequently removed from the dataset.

A new population of LDA-QSARs was developed on the training set chemicals, strictly following the procedure described in Section 4.2.3.1. The splitting of the original dataset, after removal of the outliers, resulted in 82 chemicals in the training set and 39 in the test set. The best LDA-QSAR model was chosen from a population of the best 25 developed using four variables. This selected number of variables was justified on the basis of a flattening of the $MR_{BOOTSTRAP}$ value, as detected by the bootstrap procedure for models with five, and up to ten, variables (see Figure S1 in Supplementary Materials S2). The performance of the best model, summarized in Table 4.1, was indicative of the good performance of the LDA-QSAR in terms of fitting and robustness, and considering its external predictive ability when it was applied to the respective test set.

Table 4.1. Summary of the statistical results of the LDA-QSAR. "Random range" and "random descriptors nature" indicate the probability of coincidental relationships between the molecular descriptors and the response, using randomized descriptors within their numerical ranges (random range), or considering both their numerical ranges and their nature, i.e., discrete, continuous, binary (random descriptors nature).

	n	**ACC**	**MR**	**SN**	**SP**	**P**	**AUC**	**$MR_{BOOTSTRAP}$**	**Random Range**	**Random Descriptors Nature**	**Selected Molecular Descriptors**
Training	82	0.89	0.11	0.92	0.84	0.90	0.85	$0.32 \pm 2.7 \times 10^{-3}$	3.8×10^{-3}	4.6×10^{-3}	GATS3e, ATSC6p, GATS8m, MIC2
Test	39	0.85	0.15	0.88	0.80	0.88	0.85	-	-	-	-

Specifically, the AUC values quantified for the training and the test sets were both 0.85, and the global accuracy and sensitivity were close to 0.90 in both the training and the test sets. The specificity was slightly lower, but remained consistently above 0.80. Furthermore, the probability of coincidental relationship between the molecular descriptors and the response using randomized descriptors, reported in Table 4.1, was close to zero, thereby providing additional support for the quality of the model. To ensure transparency, the linear scoring equations of the split LDA-QSAR are reported in Supplementary Materials S2 (Equations (S1) and (S2)), along with the ROC plots of the model (Figures S2 and S3). The analysis of the AD (Figure S5) highlighted that the majority of the compounds fell within the AD of the model. However, in the training set, 1,6-diiodoperfluorohexane (CASRN 375-80-4) had a cos α value (0.7690) that deviated considerably from the threshold (cos α (t-95% = 0.9661)), while two other compounds (perfluorohexanoic acid, CASRN 307-24-4; 3,3-bis(trifluoromethyl)-2-propenoic acid, CASRN 1763-28-6) had cos α values (0.9372 and 0.9593, respectively) only slightly lower than the threshold. These results indicate the dissimilarity of these PFAS from the rest of the compounds within the structural space defined by the molecular descriptors selected in the model (i.e., GATS3e, ATSC6p, GATS8m, and MIC2, commented on below). In particular, 1,6-diiodoperfluorohexane was the compound with the lowest value of ATSC6p, and the second highest value of MIC2. Similarly, perfluorohexanoic acid was the compound with the highest value of GATS8m. Finally, 3,3-bis(trifluoromethyl)-2-propenoic acid was characterized by the largest value of GATS3e and by a relatively high value of ATSC6p in comparison with

the other compounds. Nevertheless, the model accurately predicted 1,6-diiodoperfluorohexane and perfluorohexanoic acid.

Three molecular descriptors out of the four selected in the model are autocorrelation descriptors, i.e., GATS3e, ATSC6p, and GATS8m. Autocorrelation descriptors represent a vast class of global 2D descriptors that have been extensively used for the development of QSAR models across diverse areas of research [66,67,68]. GATS3e and GATS8m are calculated from the Geary's autocorrelation coefficient and encode for the spatial distribution along the molecular structure of, respectively, electronegativity at lag 3, and atomic mass at lag 8 [69]. It is interesting to note that classification QSAR models developed in previous studies identified similar autocorrelation descriptors weighted by atomic masses (i.e., GATS3m, HATS6m) as relevant to discriminate between the hTTR binding activity degree of PFAS [35,36]. ATSC6p is calculated from Moreau–Broto's autocorrelation coefficient [70], and reflects the spatial distribution of polarizability at lag 6 [69]. Interestingly, a recent molecular docking analysis identified hydrogen bonds and hydrophobic interactions as the driving forces of the hTTR binding of PFAS [30], where electronegativity and polarizability are critical factors for the formation of these interactions, respectively. These findings supported the selection of descriptors, such as GATS3e and ATSC6p, which encode these types of electronic properties. Furthermore, the use of autocorrelation descriptors may provide information regarding the length and the configuration of molecular structures as they use topological distances to represent the distances between atoms. The descriptor MIC2 is defined as the modified information content index of the neighborhood symmetry of 2-order, and belongs to the information content descriptor class [69]. Information content descriptors measure the degree of diversity within a molecule and therefore are used to describe its complexity. As described by King and colleagues [71], the MIC indices weigh the individual terms of the information content by the atomic weight of the constituent atoms of a compound. Hence, the resulting values encode for the molecular complexity by taking into account the atomic weight of the constituent atoms. Previous literature work identified the descriptor IC3 (i.e., the information content index of neighborhood symmetry of 3-order, similar to MIC2) as relevant to predict the hTTR binding of PFAS [34]. The selected molecular descriptors are comprehensively described in Supplementary Materials S2 (Table S16).

Only a limited number of compounds were misclassified (nine compounds in the training set and five compounds in the test set), thereby demonstrating the predictive ability of the model. Potential sources of misclassification are discussed as follows. The

results of the PCA performed on the molecular descriptors of the model (see Table S5 in Supplementary Materials S1) suggested that the misclassified compounds assigned a priori to a specific class actually fell within the structural space dominated by compounds associated with the other class. The PCA did not reveal any other relevant structural patterns or clusters of misclassified compounds. Therefore, the calculation of similarity values using the Euclidean distance on the molecular descriptors of the model was performed (see Table S6 in Supplementary Materials S1), which confirmed the PCA results. The errors were therefore attributed to a high structural similarity of the misclassified compounds with those belonging a priori to the opposite class, which were instead correctly classified. One possible explanation for these errors is that the molecular descriptors selected in the model are not sufficiently sensitive to small variations in the molecular structure of similar compounds that belong to opposite a priori classes. In other cases, compounds were misclassified due to their post probability values being close to 0.5 (e.g., (perfluorobutyryl)-2-thenoylmethane, CASRN 559-94-4; perfluorobutanoic acid, CASRN 375-22-4; 1H,1H,9H-Perfluorononyl acrylate, CASRN 4180-26-1), or due to their median activity values being close to 50% (e.g., octafluoroadipamide, CASRN 355-66-8; 1H,1H,9H-Perfluorononyl acrylate, CASRN 4180-26-1). Another possible explanation for misclassifications could be the presence of mistakes in the experimental measures used to generate the models. Nevertheless, the proposed LDA-QSAR was likely to be precautionary, which is a favorable attribute in QSARs developed for hazards predictions, since the sensitivity quantified for the training and the test sets were both greater than specificity, and a greater proportion of PFAS were misclassified as active than those misclassified as weak/not active. Furthermore, the post probabilities of PFAS misclassified as weak/not active were found to be closer to 0.5 than those of PFAS misclassified as active.

After the external validation of the model, the training and test sets were pooled together and the entire dataset was used to recalculate the model, thereby capturing all the experimental and structural information included in the full dataset, composed of 121 compounds. The linear scoring equations of the full LDA-QSAR for the two classes are reported below:

$$\boldsymbol{Class\ A\ score} = -47 + 37 \cdot \boldsymbol{GATS3e} + 79 \cdot \boldsymbol{ATSC6p} + 10 \cdot \boldsymbol{GATS8m} + 39 \cdot \boldsymbol{MIC2} + log\,(0.61) \quad (4.1)$$

$$\boldsymbol{Class\ I\ score} = -38 + 30 \cdot \boldsymbol{GATS3e} + 73 \cdot \boldsymbol{ATSC6p} + 8.1 \cdot \boldsymbol{GATS8m} + 32 \cdot \boldsymbol{MIC2} + log\,(0.39) \quad (4.2)$$

(n° Training set = 121; ACC = 0.87; MR = 0.13; SN = 0.95; SP = 0.74; P = 0.85; AUC = 0.85)

The here-proposed models should be applied to discriminate between active and weak/not active compounds belonging to the PFAS group. The Shannon entropy values are reported in Supplementary Materials S1 (Table S7). The ROC plot and the AD plot regarding the LDA-QSAR developed on the full dataset are reported in Supplementary Materials S2 (Figures S4 and S6, respectively).

4.3.2 MLR-QSARs

A preliminary modelling attempt was performed using all 68 compounds included in the regression dataset. This showed the presence of five outliers, which adversely impacted QSARs performances: perfluorooctanesulfonyl fluoride (CASRN 307-35-7), 6:1 fluorotelomer alcohol (CASRN 375-82-6), 1H,1H,8H,8H-perfluoro-3,6-dioxaoctane-1,8-diol (CASRN 129301-42-4), 4H-perfluorobutanoic acid (CASRN 679-12-9), and perfluoro-1,4-diiodobutane (CASRN 375-50-8). Perfluorooctanesulfonyl fluoride is the only sulfonyl fluoride included in the dataset, whose Log RP value is −2.5. This value is considerably lower than the Log RP values for the six compounds in the dataset that are most similar to perfluorooctanesulfonyl fluoride, as identified using PCA (i.e., min Log RP = −0.93; max Log RP = −0.032; mean Log RP = −0.38; median Log RP = −0.35). Similarly, 6:1 fluorotelomer alcohol has the lowest Log RP value (equal to −2.7) among all the other fluorotelomers included in the dataset (i.e., min Log RP = −1.98; max Log RP = −1; mean Log RP = −1.5; median Log RP = −1.5). In a similar manner, 1H,1H,8H,8H-perfluoro-3,6-dioxaoctane-1,8-diol is the compound with the lowest Log RP value (equal to −2.8) among all the per- or polyfluoroethers included in the dataset (i.e., min Log RP = −2.3; max Log RP = −0.0082; mean Log RP = −0.90; median Log RP = −0.69). Comparable results were obtained for compounds with at least one ether bond included in the structure, regardless of the fluorination degree of the carbon atoms involved in that bond (i.e., min Log RP = −2.3; max Log RP = −0.0082; mean Log RP = −0.94; median Log RP = −0.68).

The five mentioned outliers were removed from the dataset. The subsequent splitting procedure resulted in 43 chemicals in the training set and 20 chemicals in the test set, and a new population of MLR-QSARs was developed on the training set chemicals,

following the procedure described in Section 4.2.3.2. The best MLR-QSAR model was chosen from a population of the best 25 developed using three variables. Specifically, the MAE was quantified for different populations of bootstrapped models, at increasing levels of complexity, ranging from one up to six variables. The results of this analysis indicated a progressive increase in $MAE_{BOOTSTRAP}$ values in models with more than three variables (see Figure S7 in the Supplementary Materials S2). The fitting and the robustness of the models, which was checked by leave-one-out cross-validation and randomization of the descriptors, were evaluated using several metrics. These metrics, with the additional support of regression diagnostic plots, were then used to select the optimal QSAR among the population of available models with up to three molecular descriptors (25 models for each variable size). The equation of this best split model based on three molecular descriptors (i.e., piPC5, GGI9, and AATSC0e, commented on below) and the plot of experimental versus predicted response values are reported in the Supplementary Materials (see Supplementary Material S2—Equation (S3), and Figure S8), while the values of the statistical metrics are summarized in Table 4.2. In Equation (S3), the descriptors are listed in order of relative importance according to their standardized regression coefficients. Standardized regression coefficients of the descriptors are reported in Table 4.3.

Table 4.2. Summary of the statistical results of the MLR-QSAR. "Random range" and "random descriptors nature" indicate the probability of coincidental relationships between the molecular descriptors and the response, using randomized descriptors as defined in Table 4.1.

	n	**R^2**	**MAE**	**Q^2_{loo}**	**Q^2_{F3}**	**R^2_{YS}**	**$MAE_{BOOTSTRAP}$**	**Random Range**	**Random Descriptors Nature**	**Selected Molecular Descriptors**
Training	43	0.81	0.30	0.77	-	0.072	$0.58 \pm 5.7 \times 10^{-3}$	8.3×10^{-10}	3.3×10^{-10}	piPC5, GGI9, AATSC0e
Test	20	0.77	0.26	-	0.82	-	-	-	-	-

The values of R^2 and Q^2_{LOO}, which were around 0.80, confirmed the good performance of the MLR-QSAR in terms of its ability to fit the data and of its internal robustness. Moreover, the absence of chance correlation was confirmed by the probability of coincidental relationships and an R^2_{YS} close to zero. The calculation of small and consistent values of MAE, both for the training and for the external test set,

demonstrated the internal and external predictivity of the model. These results were consistent with the values of R^2 and Q^2_{F3}, quantified for the test set, which were close to or above 0.80, respectively. Finally, the ratio of the number of training set data to the number of molecular descriptors was 14.3, which significantly exceeded the minimum threshold of 5 that is commonly used to control the risk of chance correlations in a QSAR model [43].

An investigation of the regression diagnostic plots (reported in Supplementary Materials S2) did not highlight any relevant anomalies. All the data points were regularly distributed along the diagonal of the experimental versus predicted response values (Figure S8), while the plot of the residuals confirmed the homoscedasticity of the residuals along the range of the predicted values (Figure S10). The Williams plot (Figure S12) showed, on the y-axis, that the standardized residuals of all the predictions fell within ±2.5 standard deviation units, indicating accurate Log RP predictions. On the x-axis, only two compounds (perfluorobutanoic acid, CASRN 375-22-4; perfluorotetradecanoic acid, CASRN 376-06-7) had leverage values (h = 0.2996, and h = 0.3732, respectively) slightly larger than the cut-off h^* (0.2791), which underlined their distance from the centroid of the model, defined by the molecular descriptors selected in the model. Indeed, perfluorobutanoic acid and perfluorotetradecanoic acid were the two compounds with the lowest and the greatest values of the descriptors piPC5 and of GGI9, respectively. The most important descriptor selected in the model, according to the related standardized coefficient, was piPC5, which is a topological descriptor defined as a conventional bond order ID number of order 5 that belongs to the path count descriptor group [69]. The positive sign in Equation (S3) suggested that the T4-hTTR competing potency was positively related to conventional bond order. piPC5 provides information about the length, the form [72], and the linear structure of a compound [73]. As indicated by Jia et al. [73], compounds with multiple linear structures exhibit greater values of piPC5, which is linked to higher hydrophobicity. Therefore, the positive sign of piPC5 in Equation (S3) was consistent with the findings of previous molecular docking analysis that identified hydrophobic interactions as drivers for hTTR binding of PFAS [30], which is justified by the hydrophobic nature of the hTTR binding site for T4 [15]. Furthermore, piPC5 values of chemicals in the training set had a strong positive correlation (0.78) with the chain length, which was recognized in previous molecular docking and QSAR studies as an additional driving structural feature for the hTTR binding of PFAS [30,34,35,74]. In a study by Kovarich et al. [35], the average molecular weight (AMW) of PFAS was also identified as a relevant property to discriminate between hTTR binders and non-

hTTR binders. In this study, a strong positive correlation (0.82) was observed between piPC5 and AMW, which suggested again the possible role of the molecular dimension in determining the strength of T4-hTTR competitors. The second most significant molecular descriptor, based on its standardized coefficient, was GGI9, which exhibited a negative correlation with Log RP. GGI9 belongs to the family of topological charge descriptors [69,75]. Topological charge indices evaluate the charge transfer between pairs of atoms, and therefore the global charge transfer, in a molecule. GGI9 is defined as the topological charge index of order 9, thus it encodes for the total charge transfer between atoms placed at a topological distance of 9. Topological charge indices are associated with the molecular dipole moment, encoding information about the potential polar interactions that may contribute to chemical behaviors, such as lipophilicity [30,69]. The topological distance encoded by GGI9 provided further information regarding the length and the configuration of molecular structures influencing the binding to hTTR. In this study, the GGI9 values ranged from 0.04 to 0.74 for 21 out of 43 chemicals in the training set, while for the remaining PFAS, the GGI9 value was 0. Interestingly, when chemicals exhibited GGI9 values of 0, the piPC5 values were lower (from 2.3 to 3.8) compared to the piPC5 values of the other chemicals in the training set (which ranged from 3.6 to 4.6). Therefore, GGI9 provided additional information for differentiating the T4-hTTR competing potency of the most active PFAS (i.e., those with larger values of piPC5), depending on their length and configuration. Finally, AATSC0e was the least influential descriptor in Equation (S3) based on its standardized coefficient, and it is inversely correlated with the response. AATSC0e is an averaged, centered, autocorrelation descriptor calculated from Moreau–Broto's autocorrelation coefficient [70]. It reflects the spatial distribution of electronegativity along the structure of a compound at lag 0. As previously discussed, former molecular docking analysis identified that hydrogen bonds drive hTTR binding of PFAS [30], where electronegativity plays a key role. This finding is consistent with the observation that compounds in the training set with low values of AATSC0e also exhibited large experimental values of Log RPs. The definitions and the correlations between the molecular descriptors selected in the regression model are reported in Supplementary Materials S2 (Tables S17–S19). In conclusion, in order to use all the experimental and structural information included in both the training and the test sets, the MLR-QSAR was recalibrated on the full dataset (63 chemicals in total). The equation of the full model is reported below, together with metrics for the evaluation of the fitting and the robustness of the model:

$$\boldsymbol{LogRP} = -2.7\ (\pm 1.3) + 1.6\ (\pm 0.26) \cdot \boldsymbol{piPC5} - 3.3\ (\pm 0.79) \cdot \boldsymbol{GGI9} - 11\ (\pm 2.7) \cdot \boldsymbol{AATSC0e} \quad (4.3)$$

(n° Training set = 63; R^2 = 0.80; MAE_{TR} = 0.28; Q^2_{loo} = 0.77; R^2_{YS} = 0.049).

Standardized regression coefficients of the descriptors in Eq. 4.3 are reported in Table 4.3.

Table 4.3. List of descriptors selected in Eq. (S3) and Eq. 4.3 and their standardized regression coefficients.

Descriptor	Standardized regression coefficient	
	Split model (Eq. (S3))	Full model (Eq. 4.3)
piPC5	0.9721	1.0013
GGI9	− 0.6698	− 0.6618
AATSC0e	− 0.472	− 0.4839

Equation 4.3 is proposed as the MLR-QSAR model to predict the T4-hTTR competing potency (expressed as Log RP) of new PFAS. As expected, the model has analogous coefficients to those observed in Equation (S3), and consistent values of the statistical metrics and applicability domain. The prediction uncertainty values are reported in Supplementary Materials S1 (Table S8), and diagnostic plots regarding the full MLR-QSAR are reported in Supplementary Materials S2 (Figures S9, S11 and S13).

4.3.3 Case study: screening the potential hTTR disruption by PFAS included in the OECD List

In this section, the screening of the hTTR disruption of the PFAS included in the OECD List was addressed as a case study to demonstrate the application of the here-proposed LDA- and MLR-QSARs in a sequential approach. First, the LDA-QSAR was applied to discriminate active from weak/not active compounds. Second, the MLR-QSAR was used to provide a quantitative estimation of the T4-hTTR competing potency (expressed as Log RP) of the active PFAS. In order to facilitate this procedure, and to make it available to scientists interested in the estimation of the hazardous properties of PFAS from the molecular structure, the QSARs proposed in this study were implemented in the QSAR-ME Profiler beta version 1.02, a non-commercial software freely available online (https://dunant.dista.uninsubria.it/qsar/). The models were

applied according to the aforementioned approach to screen the 2934 neutral organic PFAS remaining in the OECD List after the data curation procedure explained in Section 4.2.5. The original names of the structural categories to which the PFAS belong, as provided in the OECD List, were used in this study to support the analysis and are reported from now on between quotation marks ("").

The LDA-QSAR was initially applied to discriminate between active and weak/non active PFAS. Predictions falling outside the AD of the model, and/or with a post probability lower than 0.75, were deemed unreliable and excluded from further analysis.

Table 4.4. The coverage (AD) of the curated OECD List in the LDA-QSAR and in the MLR-QSAR, following the sequential approach, for each structural category.

	LDA-QSAR			**MLR-QSAR**		
Structure Category	**Total (%)**	**Inside AD**		**Total (%)**	**Inside AD**	
-	**-**	**Number of PFAS (%)**	**Number of Structural Subcategories**	**-**	**Number of PFAS (%)**	**Number of Structural Subcategories**
Fluorotelomer—related compounds	1086 (37.0)	436 (40.1)	24	214 (31.5)	147 (68.7)	19
Other PFAA precursors or related compounds—semifluorinated	686 (23.4)	279 (40.7)	8	62 (9.1)	16 (25.8)	6
Perfluoroalkyl carbonyl compounds	359 (12.2)	156 (43.5)	9	79 (11.6)	56 (70.9)	6
Per- and polyfluoroalkyl ether-based compounds	280 (9.6)	119 (42.5)	18	95 (14.0)	69 (72.6)	15
Perfluoroalkane sulfonyl compounds	271 (9.2)	124 (45.8)	9	75 (11.0)	44 (58.7)	7
Other PFAA precursors and related compounds—perfluoroalkyl ones	240 (8.2)	161 (67.1)	10	147 (21.6)	81 (55.1)	10
Perfluoroalkyl phosphate compounds	12 (0.4)	8 (66.7)	2	8 (1.2)	5 (62.5)	2
Total	2934 (100)	1283 (43.7)	-	680 (100)	418 (61.5)	-

As is reported in Table 4.4, nearly 40% of the predictions were reliable. The structural categories "other PFAA precursors and related compounds—perfluoroalkyl ones" and "fluorotelomer-related compounds" were the most and the least covered by the AD of

the LDA-QSAR, respectively. A summary of the AD coverage of the LDA-QSAR, focused on each structural category and described by cause, is reported in Table S9. Interestingly, according to these results, the majority of the PFAS were excluded because of post probability falling below the threshold. This was particularly evident for the "fluorotelomer-related compounds" and the "perfluoroalkane sulfonyl compounds". Nearly one-third of the PFAS were excluded because they were outside of the structural AD. The category "Other PFAA precursors or related compounds—semifluorinated" was the least represented in the training set, given the large percentage of PFAS falling outside the structural AD. The AD coverage of the LDA-QSAR for each studied PFAS is reported in Table S4.

A total of 680 PFAS (53%) belonging to seven structural categories were predicted as active. As illustrated in Figure 4.1, the structural categories "perfluoroalkyl phosphate compounds", "other perfluoroalkyl acids (PFAA) precursors and related compounds—perfluoroalkyl ones", and "per-and polyfluoroalkyl ether-based compounds" were of major concern due to the high percentage of active predictions. The structural categories "perfluoroalkane sulfonyl", "perfluoroalkyl carbonyl", and "fluorotelomer-related compounds", although showing lower percentages of active predictions, were still of concern given the large number of PFAS belonging to them. Finally, the structural category "other PFAA precursors or related compounds—semifluorinated" was the one of least concern due to the great percentage of weak/not active predictions. Additional details about the proportions of active and weak/not active PFAS, within each structural category and subcategory in which the PFAS were further categorized, as provided in the OECD List, are summarized in Supplementary Materials S1 (Table S10) and more exhaustively described in Supplementary Materials S2. Different properties, encoded by the selected molecular descriptors, were identified as the major drivers of PFAS binding to hTTR (i.e., hydrophobicity, chain length, molecular weight, and electronegativity) and were investigated versus activity profiles across different structural categories. The aim of this analysis was to evaluate whether and how these properties are related to the differences in prediction among various structural categories. The full list of PFAS, along with the abovementioned properties, and a comparative analysis, are reported in Supplementary Materials S1 (Table S11 and Table S12, respectively). As expected, based on the mechanistic interpretation of the selected molecular descriptors, the structural category "other PFAA precursors or related compounds—semifluorinated" was the one characterized by the lowest median values of all the properties under consideration, compared to the other structural categories. On the contrary, the structural categories "perfluoroalkyl

phosphate compounds", "other perfluoroalkyl acids (PFAA) precursors and related compounds—perfluoroalkyl ones", and "per- and polyfluoroalkyl ether-based compounds" were characterized by high median values of all the properties under consideration, with few exceptions, compared to the other structural categories. Nevertheless, it is important to highlight that the activity is led by the concurrent combination of the molecular descriptors selected in a model.

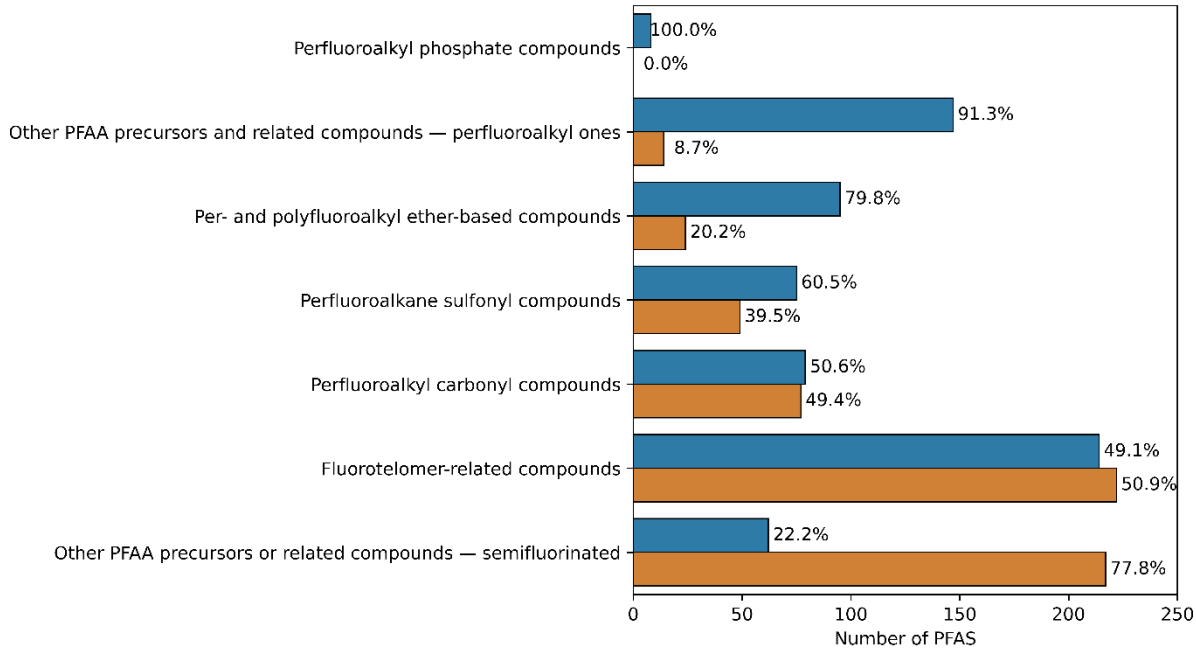

Figure 4.1. Barplot summarizing the LDA-QSAR screening results for the curated OECD List of PFAS. Blue bars indicate the number of the PFAS predicted as active; orange bars indicate the number of the PFAS predicted as weak/not active. Percentage values indicate the proportion of active or weak/not active PFAS within each structural category.

Following the sequential approach, the MLR-QSAR was then applied to quantitatively predict the T4-hTTR competing potency of the 680 active PFAS. As was mentioned in Section 4.2.5, the value Log RP ≥ −1.26 suggested in the literature [52] was used as a threshold to identify strong hTTR binders among the active PFAS. As reported in Table 4.4, after AD assessment, nearly 60% of the predictions were considered reliable. The structural category "other PFAA precursors or related compounds—semifluorinated" was the category least covered by the AD of the MLR-QSAR. With the exception of this structural category, the here-proposed MLR-QSAR was adequately sensitive toward most of the structural features belonging to the different structural categories. A

summary of the AD coverage of the MLR-QSAR, focused on each structural category and described by cause, is reported in Table S13, while the AD coverage for each studied PFAS is reported in Table S4.

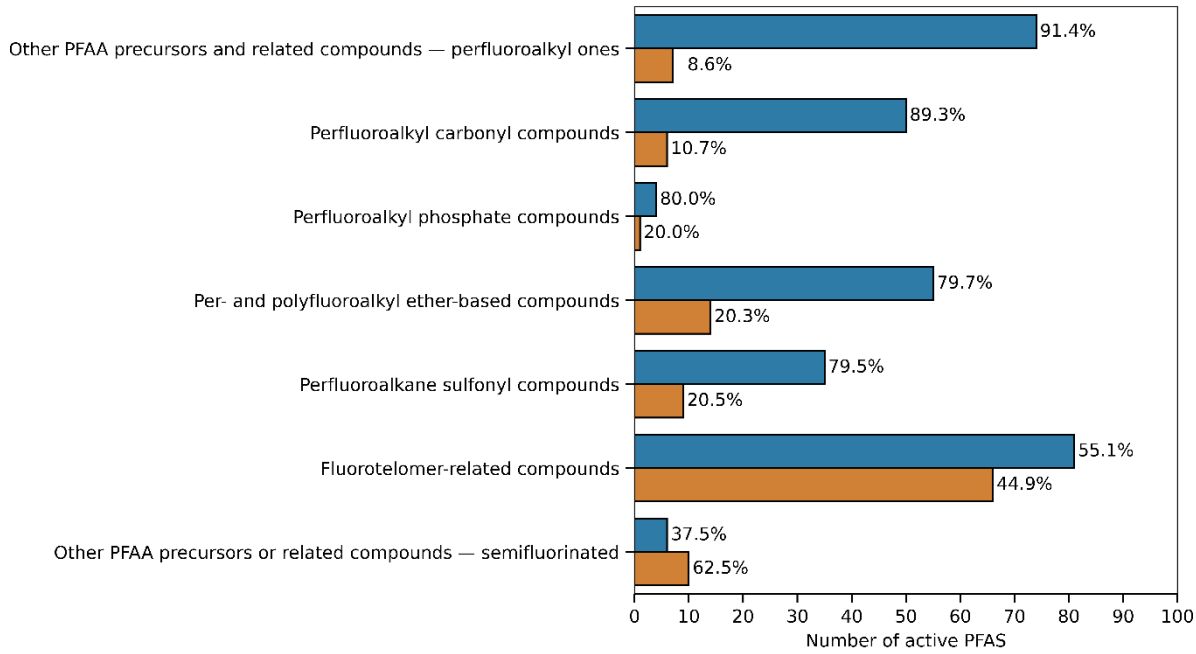

Figure 4.2. Barplot summarizing the MLR-QSAR screening results for the active PFAS that fell within the AD of the LDA-QSAR. Blue bars indicate the number of the active PFAS predicted with a Log RP ≥ −1.26 (i.e., strong hTTR binders); orange bars indicate the number of the active PFAS predicted with a Log RP < −1.26. Percentage values indicate the proportion of PFAS predicted with a Log RP ≥ −1.26 or <−1.26 within each structural category.

A total of 305 active PFAS (73%) belonging to seven structural categories were predicted as strong hTTR binders. As illustrated in Figure 4.2, the structural categories "other perfluoroalkyl acids (PFAA) precursors and related compounds— perfluoroalkyl ones", "perfluoroalkyl carbonyl", "perfluoroalkyl phosphate compounds", "per- and polyfluoroalkyl ether-based compounds", and "perfluoroalkane sulfonyl" were the structural categories of greatest concern due to percentages of strong hTTR binders roughly equal to 80% or more. The structural category "fluorotelomer-related compounds", although showing a lower proportion of strong hTTR binders, still remained of relative concern due to the large number of PFAS in the category. Finally, the structural category "other PFAA precursors or

related compounds—semifluorinated" was the only one predominantly characterized by PFAS of lower hTTR binding strength. Additional details about the proportions of strong hTTR binders among active PFAS, within each structural category and subcategory, are summarized in Supplementary Materials S1 (Table S14) and more exhaustively described in Supplementary Materials S2.

It is also significant to point out that PFAS with positive values of Log RP show a stronger binding affinity to hTTR than its natural ligand T4. Based on results reported in Table S4, a total of 49 PFAS belonging to the structural categories "perfluoroalkyl carbonyl compounds", "perfluoroalkane sulfonyl compounds", "per- and polyfluoroalkyl ether-based compounds", and "other PFAA precursors and related compounds—perfluoroalkyl ones" had positive Log RP values, further indicating a particular need for additional studies on these categories of PFAS.

To provide an additional validation of the predictions generated by the here-proposed QSARs, a thorough and extensive search was conducted for existing experimental data on hTTR disruption within the literature for the PFAS included in the OECD List, among those with reliable predictions. This led to the collection of *in vitro* measured experimental data for only 12 PFAS from six different references [36,40,76–79]. Four out of these PFAS had multiple data from different studies. The list of the 12 PFAS, along with their experimental outcomes and the corresponding literature references, is reported in Supplementary Materials S2 (Table S20). Nine out of the twelve predictions (i.e., 75%) showed full agreement with all the available experimental data. Remarkably, for three out of these nine PFAS, the predictions fully agreed with multiple data from different studies. These results supported the quality of the here-proposed QSARs. On the contrary, a disagreement between predictions and experimental data was observed for the remaining three PFAS, which were wrongly classified as active. Among these, in one case, the prediction agreed with one of the multiple studies. Nevertheless, it should be highlighted that the predictions were likely to be precautionary. Despite the limited experimental data on hTTR disruption of PFAS, the high level of agreement between them and those generated by the here-proposed QSARs is promising, providing a strong indication of their reliability. Nevertheless, while these results are encouraging, further experimental validation across a broader range of PFAS using new experimental data is desirable.

4.3.4 Comparison with previous studies

In the past, a limited number of QSAR studies to predict the potential hTTR disruption of PFAS were performed. However, most of these models used proprietary software during QSARs development, which may limit their application. The models were characterized by narrow structural and response domains because of the small sizes of the training sets, which limited their applicability to a broader range of chemical structures and responses. Furthermore, as was mentioned in the introduction, experimental data used for model development have been measured with inconsistent methodologies (i.e., different assays were used to determine the hTTR binding affinity), and/or assays currently not being validated as reliable and fit for purpose by the European Commission's EURL ECVAM.

The summary and comparison of the classification and regression QSARs developed in the present work, with those developed in previous studies, are presented in Table 4.5 and in Table 4.6, respectively. The presented comparison clearly shows that the new models exhibited similar or lower complexity (i.e., number of descriptors), as well as comparable or better performances than the previous models. Moreover, the new models were based on larger datasets, which included about two to six times the number of chemicals modelled in other studies. Consequently, the new models had larger ADs (see Supplementary Materials S1, Table S15). It is noteworthy that the number of descriptors in the new models was chosen by studying the behaviour of $MR_{BOOTSTRAP}$ values in classification and $MAE_{BOOTSTRAP}$ values in regression, with the aim of reducing the risk of overfitting. It is important to note that in the former studies, no comparable procedures were applied to ascertain that overfitting did not take place. Even if the ratio "training set size/number of descriptors" was above or equal to five in all the regression models, which is a simple measure to minimize the risk of chance correlations [43], this did not exclude the possibility of overfitting in the previous QSARs, considering the smaller dimensions of their training sets. Furthermore, unlike the other QSARs, the here-proposed models were not based on commercial descriptors, which enhances their applicability. Nevertheless, the consistency of descriptors across all the different models, even if calculated using different software and selected using different algorithms while modelling different specific endpoints, validated the significance of lipophilicity, chain length, and molecular weight as particularly relevant for the assessment of PFAS activity as hTTR disruptors.

*Table 4.5. Comparative analysis between the present LDA-QSAR and classification QSARs reported in previous research. * Four different models were presented. ** Values of SN, SP, and ACC were calculated in this study from data reported in the Supporting Information of the original paper. kNN: k-nearest neighbor; DTC: decision tree classifier. N/A stands for "not available".*

	Present Model	**Kar et al. [34]**	**Kovarich et al. * [35]**	**Sosnowska et al. ** [36]**
Endpoint	hTTR binding affinity	hTTR binding affinity	hTTR binding affinity	RPF
In vitro assay	ANSA-based [42]	RLBA [40]	RLBA [40]	TTR-TRβ CALUX [36]
Method	LDA	LDA	kNN	DTC
Dataset size	121	24	19	44
Training set size	82	16	10	33
Test set size	39	8	9	11
Number of descriptors	4	3	2	2
SN training	0.92	1	0.83–1	0.96
SN test	0.88	1	1	1
SP training	0.84	0.83	0.75–1	1
SP test	0.80	1	0.75–1	0.50
ACC training	0.89	0.94	0.90–1	0.97
ACC test	0.85	1	0.90–1	0.91
P training	0.90	0.91	N/A	1
P test	0.88	1	N/A	0.90
AUC training	0.85	0.95	N/A	N/A
AUC test	0.85	1	N/A	N/A

*Table 4.6. Comparative analysis between the present MLR-QSAR and regression QSARs reported in previous research. * Thirty-one different models were reported in Sosnowska et al. [36]; the sizes of the training and test sets were extracted from the Supporting Information of the original paper. IC50: inhibitory concentration 50%. N/A stands for "not available".*

	This Model	**Kar et al. [34]**	**Sosnowska et al. [36] Approach 1**	**Sosnowska et al. [36] Approach 2 ***
Endpoint	RP	IC50	RPF	RPF
Method	MLR	MLR	MLR	MLR
In vitro assay	ANSA-based [42]	RLBA [40]	RLBA [40]	TTR-TRβ CALUX [36]
Dataset size	63	15	35	35
Training set size	43	10	24	25
Test set size	20	5	11	10
Number of descriptors	3	2	3	4–5
Ratio training set size/number of descriptors	14.3	5	8	5–6.3
R^2	0.81	0.86	0.77	N/A
R^2_{EXT}	0.77	0.64	N/A	N/A
MAE_{TR}	0.30	N/A	0.43	N/A
MAE_{TEST}	0.26	0.11	0.40	0.34–0.54
Q^2_{loo}	0.77	0.73	0.77	0.76–0.82
Q^2_{F3}	0.82	N/A	0.81	0.76–0.82
R^2_{YS}	0.07	N/A	0.13	N/A

The implementation of the new models into dedicated, and freely available, software facilitated their application for screening purposes, with the clear quantification of their domains and of the uncertainty of predictions. It represented a clear advantage, compared to other models, to assist in the screening of larger numbers of PFAS, such as those included in the OECD List presented in the case study.

4.4 Conclusions

Only a limited number of studies are available that report data on TH activity by PFAS. Consequently, the number of previously published QSARs was limited, and they were based on commercial descriptors, and though characterized by good performances, they only had small ADs. Now, new, simple, robust, and predictive QSAR models were developed in this study to assess the capability of PFAS to bind to hTTR and disrupt hTTR function, which is a critical mode of action known to disrupt the TH system. Two QSAR models were proposed, one to identify hTTR-binding PFAS, and another to quantify their ability to compete with the thyroid hormone T4 for binding to hTTR, in terms of relative competitive potency. The new models were calibrated on larger and homogeneous datasets, including two to six times the amount of data compared to those available for previous models, including most of the chemicals used in previous studies, in addition to tens of other PFAS. Therefore, they had larger ADs and a greater ability to provide reliable predictions for a broader range of PFAS. The size of the dataset allowed for the application of rigorous statistical procedures to detect and avoid overfitting and random correlations, as well as to demonstrate the predictive ability of the QSARs. The statistical metrics calculated for the new models demonstrated their robustness and their capacity to predict the activity of PFAS that had not been used to train the models. In addition, the molecular descriptors, selected in the models by statistical procedures, were consistent with previous *in vitro* and *in silico* findings regarding the major drivers of PFAS binding to hTTR. These findings highlighted the importance of hydrogen bond formation and of hydrophobic interactions, and they pointed out the relevance of lipophilicity, molecular weight, and the chain length of molecular structures. Moreover, similar descriptors were selected in previous QSARs, which were developed using different quantitative approaches and data, thus strengthening confidence in the relevance of these descriptors to describe PFAS binding to hTTR. The utility and the applicability of the QSARs proposed in this study were demonstrated by screening about 3000 compounds included in the OECD List of PFAS. To this end, the models were implemented in the non-commercial software QSAR-ME Profiler beta version 1.02 (freely downloadable at https://dunant.dista.uninsubria.it/qsar/), allowing for the clear quantification of their domains and of the uncertainty of predictions, to further enhance the assessment of their reliability, in order to improve their confidence. The screening allowed for the identification of the PFAS of major concern for their potential hTTR disruption, which were found to belong mainly to the structural categories “per- and polyfluoroalkyl ether-based compounds”, “other PFAA precursors and related compounds—

perfluoroalkyl ones", "perfluoroalkyl carbonyl", and "perfluoroalkane sulfonyl compounds". These quantitative results pointed out both categories of PFAS and individual compounds that are of potential concern, suggesting prospects for future research efforts. Though the new models have improved predictive capacity, the screening revealed structural categories that are still poorly covered by the AD of the proposed models (e.g., "other PFAA precursors or related compounds—semifluorinated") and are thus associated with a high number of unreliable predictions. These findings highlighted the need for additional *in vitro* testing in those areas poorly covered by the AD of the models, with the aim of enhancing the quality and extending the domain of the reliable application of the existing QSARs to a greater number of PFAS. The models and predictions generated in this study addressed a critical gap in the understanding of PFAS toxicity to the functioning of the TH system. As the here-proposed QSARs were developed and validated specifically for PFAS, their application should not be extended to other chemical classes. These findings would support the current general need for NAMs development, and particularly to improve the *in silico* hazard assessment of potentially dangerous chemicals in general, but especially for PFAS due to their environmental and health impacts, and the large number of chemicals in this group. While the new QSARs demonstrated robustness and high predictive performances, which were even confirmed through the validation with experimental data, and they offer valuable simplicity, interpretability, and ease of use, future research efforts could explore the application of more complex machine learning approaches. This could offer new insights and potentially boost the predictive ability in certain structural areas, albeit at the expense of straightforward transparency and simple use. Finally, the implementation of the QSARs proposed in this work into a dedicated and non-commercial software (i.e., QSAR-ME Profiler beta version 1.02) made them available to scientists, industry, and regulatory bodies to facilitate their application and to support the assessment of unstudied and new PFAS, to identify safer alternatives, and to inform future research studies and regulatory actions, particularly for grouping strategies development and prioritization.

Acknowledgments and Funding: This research was funded by the PhD Program in Chemical and Environmental Sciences (DiSCA) at the University of Insubria; PhD scholarship awarded to Marco Evangelista.

Supporting information: Supplementary data associated with this article can be found in the online version at https://doi.org/10.3390/toxics13070590.

4.5 References

[1] Evich, M.G., Davis, M.J.B., McCord, J.P., Acrey, B., Awkerman, J.A., Knappe, D.R.U., Lindstrom, A.B., Speth, T.F., Tebes-Stevens, C., Strynar, M.J., Wang, Z., Weber, E.J., Henderson, W.M., Washington, J.W., 2022. Science 375, 512, https://doi.org/10.1126/science.abg9065.

[2] Brunn, H., Arnold, G., Körner, W., Rippen, G., Steinhäuser, K.G., Valentin, I., 2023. Environ. Sci. Eur. 35, 20, https://doi.org/10.1186/s12302-023-00721-8.

[3] De Silva, A.O., Armitage, J.M., Bruton, T.A., Dassuncao, C., Heiger-Bernays, W., Hu, X.C., Kärrman, A., Kelly, B., Ng, C., Robuck, A., Sun, M., Webster, T.F., Sunderland, E.M., 2021. Environ. Toxicol. Chem. 40, 631–657, https://doi.org/10.1002/etc.4935.

[4] Andrews, D.Q., Stoiber, T., Temkin, A.M., Naidenko, O.V., 2023. Sci. Total Environ. 901, 165939, https://doi.org/10.1016/j.scitotenv.2023.165939.

[5] Jane, L., Espartero, L., Yamada, M., Ford, J., Owens, G., Prow, T., Juhasz, A., 2022. Environ. Res. 212, 113431, https://doi.org/10.1016/j.envres.2022.113431.

[6] Zhang, J., Naveed, H., Chen, K., Chen, L., 2025. Toxics, 13, 47, https://doi.org/10.3390/toxics13010047.

[7] Coperchini, F., Croce, L., Ricci, G., Magri, F., Rotondi, M., Imbriani, M., Chiovato, L., 2021. Front. Endocrinol. 11, 612320, https://doi.org/10.3389/fendo.2020.612320.

[8] Zoeller, R.T., Tan, S.W., Tyl, R.W., 2007. Crit. Rev. Toxicol. 37, 11–53, https://doi.org/10.1080/10408440601123446.

[9] Bassett, J.H.D., Williams, G.R., 2016. Endocr. Rev. 37, 135–187, https://doi.org/10.1210/er.2015-1106.

[10] Klein, I., Ojamaa, K., 2001. N. Eng. J. Med. 344, 501–509, https://doi.org/10.1056/NEJM200102153440707.

[11] Mullur, R., Liu, Y.-Y., Brent, G.A., 2014. Physiol. Rev. 94, 355–382, https://doi.org/10.1152/physrev.00030.2013.

[12] Schroeder, A.C., Privalsky, M.L., 2014. Front. Endocrinol. 5, 40, https://doi.org/10.3389/fendo.2014.00040.

[13] Haigis, A.-C., Vergauwen, L., LaLone, C.A., Villeneuve, D.L., O'Brien, J.M., Knapen, D., 2023. Toxicol. Sci. 195, 1–27, https://doi.org/10.1093/toxsci/kfad063.

[14] Noyes, P.D., Friedman, K.P., Browne, P., Haselman, J.T., Gilbert, M.E., Hornung, M.W., Barone, S., Crofton, K.M., Laws, S.C., Stoker, T.E., Simmons, S.O., Tietge, J.E., Degitz, S.J., 2019. Environ. Health Perspect. 127, 095001, https://doi.org/10.1289/EHP5297.

[15] Gião, T., Saavedra, J., Cotrina, E., Quintana, J., Llop, J., Arsequell, G., Cardoso, I., 2020. Int. J. Mol. Sci. 21, 2075, https://doi.org/10.3390/ijms21062075.

[16] Du, X., Wu, Y., Tao, G., Xu, J., Du, Z., Wu, M., Gu, T., Xiong, J., Xiao, S., Wei, X., Ruan, Y., Xiao, P., Zhang, L., Zheng, W., 2024. Sci. Total Environ. 953, 175958, https://doi.org/10.1016/j.scitotenv.2024.175958.

[17] Escobar, G.M. de, Obregón, M.J., Rey, F.E. del, 2004. Best Pract. Res. Clin. Endocrinol. Metab. 18, 225–248, https://doi.org/10.1016/j.beem.2004.03.012.

[18] Rodrigues, V.G., Henrique, G., Sousa-Vidal, É.K., de Souza, R.M.M., Tavares, E.F.C., Mezzalira, N., Marques, T. de O., Alves, B.M., Pinto, J.A.A., Irikura, L.N.N., Silva, R.E.C. da, de Oliveira, K.C., Maciel, R.M. de B., Giannocco, G., Serrano-Nascimento, C., 2024. Endocrines 5, 430–453, https://doi.org/10.3390/endocrines5030032.

[19] Moog, N.K., Entringer, S., Heim, C., Wadhwa, P.D., Kathmann, N., Buss, C., 2017. Neuroscience 342, 68–100, https://doi.org/10.1016/j.neuroscience.2015.09.070.

[20] Coperchini, F., Teliti, M., Greco, A., Croce, L., Rotondi, M., 2024. Expert Rev. Endocrinol. Metab. 19, 307–315, https://doi.org/10.1080/17446651.2024.2351885.

[21] Bernasconi, C., Langezaal, I., Bartnicka, J., Asturiol, D., Bowe, G., Coecke, S., Kienzler, A., Liska, R., Milcamps, A., Munoz-Pineiro, M.A., Pistollato, F., Whelan, M., 2023. Publications Office of the European Union, Luxembourg, https://doi.org/10.2760/862948.

[22] Ao, J., Tang, W., Liu, X., Ao, Y., Zhang, Q., Zhang, J., 2024. J. Hazard Mater. 464, 133018, https://doi.org/10.1016/j.jhazmat.2023.133018.

[23] Cao, Y., Ng, C.A., 2025. J. Hazard Mater. 487, 137235, https://doi.org/10.1016/j.jhazmat.2025.137235.

[24] Fenton, S.E., Ducatman, A., Boobis, A., DeWitt, J.C., Lau, C., Ng, C., Smith, J.S., Roberts, S.M., 2021. Environ. Toxicol. Chem. 40, 606–630 https://doi.org/10.1002/etc.4890.

[25] European Chemicals Agency (ECHA), 2024. Key Areas of Regulatory Challenge. https://doi.org/10.2823/858284.

[26] Organisation for Economic Co-operation and Development (OECD), 2018. Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. https://doi.org/10.1787/9789264304741-1-en.

[27] Barton-Maclaren, T.S., Wade, M., Basu, N., Bayen, S., Grundy, J., Marlatt, V., Moore, R., Parent, L., Parrott, J., Grigorova, P., Pinsonnault-Cooper, J., Langlois, V.S., 2022. Environ. Res. 204, 112225, https://doi.org/10.1016/j.envres.2021.112225.

[28] Judson, R.S., Paul Friedman, K., Houck, K., Mansouri, K., Browne, P., Kleinstreuer, N.C., 2018. Curr. Opin. Toxicol. 9, 40–47, https://doi.org/10.1016/j.cotox.2018.10.002.

[29] Ramhøj, L., Axelstad, M., Baert, Y., Cañas-Portilla, A.I., Chalmel, F., Dahmen, L., De La Vieja, A., Evrard, B., Haigis, A.-C., Hamers, T., Heikamp, K., Holbech, H., Iglesias-Hernandez, P., Knapen, D., Marchandise, L., Morthorst, J.E., Nikolov, N.G., Nissen, A.C.V.E., Oelgeschlaeger, M., Renko, K., Rogiers, V., Schüürmann, G., Stinckens, E., Stub, M.H., Torres-Ruiz, M., Van Duursen, M., Vanhaecke, T., Vergauwen, L., Wedebye, E.B., Svingen, T., 2023. Front. Toxicol. 5, 1189303, https://doi.org/10.3389/ftox.2023.1189303.

[30] Dharpure, R., Pramanik, S., Pradhan, A., 2023. Arch. Toxicol. 97, 755–768, https://doi.org/10.1007/s00204-022-03434-8.

[31] Tiburtini, G.A., Bertarini, L., Bersani, M., Dragani, T.A., Rolando, B., Binello, A., Barge, A., Spyrakis, F., 2024. Arch. Toxicol. 98, 3035–3047, https://doi.org/10.1007/s00204-024-03797-0.

[32] Cousins, I.T., DeWitt, J.C., Glüge, J., Goldenman, G., Herzke, D., Lohmann, R., Miller, M., Ng, C.A., Scheringer, M., Vierke, L., Wang, Z., 2020. Environ. Sci. Process. Impacts 22, 1444–1460, https://doi.org/10.1039/D0EM00147C.

[33] Guelfo, J.L., Korzeniowski, S., Mills, M.A., Anderson, J., Anderson, R.H., Arblaster, J.A., Conder, J.M., Cousins, I.T., Dasu, K., Henry, B.J., Lee, L.S., Liu, J., McKenzie, E.R., Willey, J., 2021. Environ. Toxicol. Chem. 40, 3234–3260, https://doi.org/10.1002/etc.5182.

[34] Kar, S., Sepúlveda, M.S., Roy, K., Leszczynski, J., 2017. Chemosphere 184, 514–523, https://doi.org/10.1016/j.chemosphere.2017.06.024.

[35] Kovarich, S., Papa, E., Li, J., Gramatica, P., 2012. SAR QSAR Environ. Res. 23, 207–220, https://doi.org/10.1080/1062936X.2012.657235.

[36] Sosnowska, A., Mudlaff, M., Mombelli, E., Behnisch, P., Zdybel, S., Besselink, H., Kuckelkorn, J., Bulawska, N., Kepka, K., Kowalska, D., Brouwer, A., Puzyn, T., 2025. J. Hazard Mater. 491, 137949, https://doi.org/10.1016/j.jhazmat.2025.137949.

[37] HyperChem, Hypercube, Inc.: 1115 NW 4th Street, Gainesville, Florida 32601 USA.

[38] Dragon, version 5.5 (2007) and version 6.0 (2011),

[39] Mauri, A., 2020. alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. Ecotoxicological QSARs, 801–820, https://doi.org/10.1007/978-1-0716-0150-1_32.

[40] Weiss, J.M., Andersson, P.L., Lamoree, M.H., Leonards, P.E.G., van Leeuwen, S.P.J., Hamers, T., 2009. Toxicol. Sci. 109, 206–216, https://doi.org/10.1093/toxsci/kfp055.

[41] Evangelista, M., Chirico, N., Papa, E., 2024. J. Hazard Mater. 480, 136188, https://doi.org/10.1016/j.jhazmat.2024.136188.

[42] Degitz, S.J., Olker, J.H., Denny, J.S., Degoey, P.P., Hartig, P.C., Cardon, M.C., Eytcheson, S.A., Haselman, J.T., Mayasich, S.A., Hornung, M.W., 2024. Toxicol. in Vitro 95, 105762, https://doi.org/10.1016/j.tiv.2023.105762.

[43] Organisation for Economic Co-operation and Development (OECD), 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en.

[44] Yang, X., Ou, W., Xi, Y., Chen, J., Liu, H., 2019. Environ. Sci. Technol. 53, 7019–7028, https://doi.org/10.1021/acs.est.9b00218.

[45] Yap, C.W., 2011. J. Comput. Chem. 32, 1466–1474, https://doi.org/10.1002/jcc.21707.

[46] Organisation for Economic Co-operation and Development (OECD), 2018. Summary report on the New Comprehensive Global Database of Per- and Polyfluoroalkyl substances (PFAS). https://doi.org/10.1787/1a14ad6c-en.

[47] Kwon, H., Ali, Z.A., Wong, B.M., 2023. Environ. Sci. Technol. Lett. 10, 1017–1022, https://doi.org/10.1021/acs.estlett.2c00530.

[48] Fourches, D., Muratov, E., Tropsha, A., 2010. J. Chem. Inf. Model. 50, 1189–1204, https://doi.org/10.1021/ci100176x.

[49] Papa, E., Kovarich, S., Gramatica, P., 2013. SAR QSAR Environ. Res. 24, 333–349, https://doi.org/10.1080/1062936X.2013.773374.

[50] Xi, Y., Yang, X., Zhang, H., Liu, H., Watson, P., Yang, F., 2020. Chemosphere 242, 125135, https://doi.org/10.1016/j.chemosphere.2019.125135.

[51] Yang, X., Ou, W., Zhao, S., Wang, L., Chen, J., Kusko, R., Hong, H., Liu, H., 2021. Chemosphere 280, 130627, https://doi.org/10.1016/j.chemosphere.2021.130627.

[52] Yang, X., Ou, W., Zhao, S., Xi, Y., Wang, L., Liu, H., 2021. ACS Sustain. Chem. Eng. 9, 5661–5672, https://doi.org/10.1021/acssuschemeng.1c00680.

[53] Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. J. Cheminform. 9, 61, https://doi.org/10.1186/s13321-017-0247-6.

[54] O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. J. Cheminform. 3, 33, https://doi.org/10.1186/1758-2946-3-33.

[55] Chirico, N., McLachlan, M.S., Li, Z., Papa, E., 2024. Environ. Sci. Process. Impacts 26, 400-410, https://doi.org/10.1039/D3EM00267E.

[56] Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. J. Comput. Chem. 34, 2121–2132, https://doi.org/10.1002/jcc.23361.

[57] Wold, S., Esbensen, K., Geladi, P., 1987. Chemometr. Intell. Lab. 2, 37–52, https://doi.org/10.1016/0169-7439(87)80084-9.

[58] Rücker, C., Rücker, G., Meringer, M., 2007. J. Chem. Inf. Model. 47, 2345–2357, https://doi.org/10.1021/ci700157b.

[59] Cawley, G.C., Talbot, N.L.C., 2010. J. Mach. Learn. Res. 11, 2079–2107, https://www.jmlr.org/papers/v11/cawley10a.html (last accessed on 06 October 2025).

[60] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second edition, Springer: New York, NY, USA, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[61] Todeschini, R., Ballabio, D., Grisoni, F., 2016. J. Chem. Inf. Model. 56, 1905–1913, https://doi.org/10.1021/acs.jcim.6b00277.

[62] Klingspohn, W., Mathea, M., ter Laak, A., Heinrich, N., Baumann, K., 2017. J. Cheminform. 9, 44, https://doi.org/10.1186/s13321-017-0230-2.

[63] Sushko, I., Novotarskyi, S., Körner, R., Pandey, A.K., Cherkasov, A., Li, J., Gramatica, P., Hansen, K., Schroeter, T., Müller, K.-R., Xi, L., Liu, H., Yao, X., Öberg, T., Hormozdiari, F., Dao, P., Sahinalp, C., Todeschini, R., Polishchuk, P., Artemenko, A., Kuz'min, V., Martin, T.M., Young, D.M., Fourches, D., Muratov, E., Tropsha, A., Baskin, I., Horvath, D., Marcou, G., Muller, C., Varnek, A., Prokopenko, V.V., Tetko, I.V., 2010. J. Chem. Inf. Model. 50, 2094–2111, https://doi.org/10.1021/ci100253r.

[64] Mathea, M., Klingspohn, W., Baumann, K., 2016. Mol. Inform. 35, 160–180, https://doi.org/10.1002/minf.201501019.

[65] Organisation for Economic Co-operation and Development (OECD), 2007. Lists of PFOS, PFAS, PFOA, PFCA, Related Compounds and Chemicals That May Degrade to PFCA. https://one.oecd.org/document/ENV/JM/MONO(2006)15/en/pdf (last accessed on 06 October 2025).

[66] Emonts, J., Buyel, J.F., 2023. Comput. Struct. Biotechnol. J. 21, 3234–3247, https://doi.org/10.1016/j.csbj.2023.05.022.

[67] Li, F., Wang, P., Fan, T., Zhang, N., Zhao, L., Zhong, R., Sun, G., 2024. J. Hazard Mater. 465, 133410, https://doi.org/10.1016/j.jhazmat.2023.133410.

[68] Speck-Planche, A., Kleandrova, V.V., Luan, F., Cordeiro, M.N.D.S., 2012. Bioorg. Med. Chem. 20, 4848–4855, https://doi.org/10.1016/j.bmc.2012.05.071.

[69] Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics. Wiley, https://doi.org/10.1002/9783527628766.ch20.

[70] Moreau, G., Broto, P., 1980. Nouv. J. Chim. 4, 359–360.

[71] King, J.W., 1989. Int. J. Quantum Chem. 36, 165–170, https://doi.org/10.1002/qua.560360712.

[72] Weinebeck, A., Kaminski, S., Murrenhoff, H., Leonhard, K., 2017. Tribol. Int. 115, 274–284, https://doi.org/10.1016/j.triboint.2017.05.005.

[73] Jia, T., Liu, W., Keller, A.A., Gao, L., Xu, X., Wu, W., Wang, X., Yu, Y., Zhao, G., Li, B., Deng, J., Mao, T., Chen, C., 2024. Sci. Total Environ. 957, 177835, https://doi.org/10.1016/j.scitotenv.2024.177835.

[74] Ren, X.-M., Qin, W.-P., Cao, L.-Y., Zhang, J., Yang, Y., Wan, B., Guo, L.-H., 2016. Toxicol. 366–367, 32–42, https://doi.org/10.1016/j.tox.2016.08.011.

[75] Gálvez, J., Garcia, R., Salabert, M.T., Soler, R., 1994. J. Chem. Inf. Comput. Sci. 34, 520–525, https://doi.org/10.1021/ci00019a008.

[76] Carlier, M.P., Cenijn, P.H., Baygildiev, T., Irwan, J., Escher, S.E., van Duursen, M.B.M., Hamers, T., 2024. Toxicol. Sci. 202, 250–264, https://doi.org/10.1093/toxsci/kfae131.

[77] Langberg, H.A., Choyke, S., Hale, S.E., Koekkoek, J., Cenijn, P.H., Lamoree, M.H., Rundberget, T., Jartun, M., Breedveld, G.D., Jenssen, B.M., Higgins, C.P., Hamers, T., 2024. Environ. Toxicol. Chem. 43, 245–258, https://doi.org/10.1002/etc.5777.

[78] Zhang, J., Kamstra, J.H., Ghorbanzadeh, M., Weiss, J.M., Hamers, T., Andersson, P.L., 2015. Environ. Sci. Technol. 49, 10099–10107, https://doi.org/10.1021/acs.est.5b01742.

[79] Xin, Y., Ren, X.-M., Ruan, T., Li, C.-H., Guo, L.-H., Jiang, G., 2018. Environ. Sci. Technol. 52, 9412–9418, https://doi.org/10.1021/acs.est.8b01494.

CHAPTER 5

A Review of Quantitative Structure–Activity Relationship (QSAR) Models to Predict Thyroid Hormone System Disruption by Chemical Substances

Marco Evangelista, and Ester Papa

Published in Toxics, Volume 13, 799, 2025
https://doi.org/10.3390/toxics13090799

Abstract

Thyroid hormone (TH) system disruption by chemicals poses a significant concern due to the key role the TH system plays in essential body functions, including the metabolism, growth, and brain development. Animal-based testing methods are resource-demanding and raise ethical issues. Thus, there is a recognised need for new approach methodologies, such as quantitative structure–activity relationship (QSAR) models, to advance chemical hazard assessments. This review, covering the scientific literature from 2010 to 2024, aimed to map the current landscape of QSAR model development for predicting TH system disruption. The focus was placed on QSARs that address molecular initiating events within the adverse outcome pathway for TH system disruption. A total of thirty papers presenting eighty-six different QSARs were selected based on predefined criteria. A discussion on the endpoints and chemical classes modelled, data sources, modelling approaches, and the molecular descriptors selected, including their mechanistic interpretations, was provided. By serving as a "state-of-the-art" of the field, existing models and gaps were identified and highlighted. This review can be used to inform future research studies aimed at advancing the assessment of TH system disruption by chemicals without relying on animal-based testing, highlighting areas that require additional research.

5.1 Introduction

The endocrine system is a network of glands and organs responsible for the proper production and homeostasis of hormones that control and regulate essential physiological processes, including growth, metabolism, and reproduction [1,2]. While the proper function of this intricate network is vital for maintaining hormonal homeostasis, the endocrine system is vulnerable to exogenous chemical substances known as endocrine disrupting chemicals (EDCs) [1,2]. By mimicking or blocking hormone activity, EDCs cause a wide range of severe adverse health outcomes in living organisms, including, among others, cancers and infertility [1,3,4]. This breadth and severity of effects have made exposure to EDCs a global concern for ecosystems and human health [4–6].

In mammals, three major axes characterise the endocrine system: the hypothalamic–pituitary–gonadal (HPG) axis, the hypothalamic–pituitary–adrenal (HPA) axis, and the hypothalamic–pituitary–thyroid (HPT) axis [2,6,7]. The HPT axis regulates the synthesis and release of specific hormones, i.e., thyroid hormones (TH), through a negative feedback loop, ensuring their homeostasis and appropriate physiological concentrations [8,9]. THs, primarily thyroxine (T4) and triiodothyronine (T3), are essential for regulating and coordinating a wide spectrum of physiological processes throughout all life cycle stages, from embryonic development to adult tissue functions. These processes include, among others, the regulation of metabolism and energy balance [10,11], and the influence on the immune, nervous, skeletal, reproductive, and cardiovascular systems [12–16]. Although proper TH activity is essential for normal physiological processes in adulthood [17], its importance is critically pronounced during gestation and early life stages, as THs play a lead role in placenta, brain, and nervous system development [18–21]. TH system-disrupting chemicals (THSDCs) are a specific subset of EDCs which target the TH system and interfere with the synthesis, secretion, distribution, and metabolism of THs and, ultimately, with their binding to nuclear TH receptors (TRs) for inhibiting or activating gene transcription [22,23]. To date, multiple chemical substances have been recognised as THSDCs, including polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), perchlorate, bisphenol A, phthalates, dioxins, pesticides, per- and polyfluoroalkyl substances (PFAS), and metals [22,24–26]. Exposure to TSHDCs can disrupt TH homeostasis, resulting in cognitive and neurobehavioral disorders [27], cancer [28], and immune, cardiovascular, and reproductive system dysfunctions [29–32].

Therefore, it is of utmost importance that THSDCs are identified without delays [23,33].

In the framework of the European Green Deal [34] and the Chemicals Strategy for Sustainability [35], the development and implementation of new approach methodologies (NAMs), including *in vitro* assays and *in silico* approaches, are heavily promoted to support the identification of EDCs and reduce the reliance on vertebrate animal testing [36–38]. The European Union (EU) is advancing this field by funding key dedicated research projects, such as the European Cluster to Improve Identification of Endocrine Disruptors (https://eurion-cluster.eu/). At the international level, the Organisation for Economic Co-operation and Development (OECD) included *in vitro* and *in silico* methodologies in the "Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals" as a relevant source of information to assess the ED properties of substances [39].

In previous years, the criteria for the determination of ED properties have been adopted under the main EU chemical regulations, such as Regulation (EU) No 528/2012 [40], Regulation (EC) No 1107/2009 [41], and Regulation (EC) No 1272/2008 [42]. Although there are minor differences in the terminology across these regulations, a chemical substance is recognised as an EDC if it meets the following three criteria: (i) it shows an adverse effect, (ii) it can alter the endocrine system through an endocrine mode of action, (iii) a plausible link between (i) and (ii) must be established. In this regard, the combined application of NAMs and the adverse outcome pathway (AOP) framework [43] has been suggested as an effective strategy [36,37,44]. Firstly, the development and application of NAMs can identify molecular initiating events (MIEs) in AOPs through which chemical substances can trigger specific endocrine modes of action and consequently lead to endocrine-related adverse effects. Secondly, a biologically plausible link between endocrine modes of action and adverse effects can emerge. This synergy gains even greater significance as it is now established that EDCs can disrupt various pathways involving hormone signalling, rather than the initial belief that their effects were solely mediated by interacting with nuclear receptors [7]. As is the case with other types of NAMs, a synergism between AOPs and quantitative structure–activity relationship (QSAR) models has been established [23,45,46]. The AOP network for TH system disruption developed by Noyes et al. [47] holds significant importance in the field, as it was used as foundational framework by the European Union Reference Laboratory for alternatives to animal testing (EURL ECVAM) to validate a suite of mechanistic *in vitro* assays for identifying THSDCs

[33,48]. Multiple MIEs have been well documented, which involved each step of the TH cycle [47]. Examples include, among others, the inhibition of thyroperoxidase (TPO), which is a critical enzyme for TH synthesis as it catalyses tyrosine residue iodination; binding to serum TH distributor proteins, such as transthyretin (TTR), thyroid binding globulin (TBG), and albumin, which serve as buffers of TH in the bloodstream to ensure the proper TH concentration in their free form; and binding to TRs, which are proteins that, once bound to TH, regulate gene expression and ultimately biological effects [47].

Despite the growing need and interest to advance TH system disruption assessments using *in silico* and QSAR approaches, a comprehensive review on this topic is currently lacking. While valuable studies have been recently published [49,50], their scopes were different. Sellami and co-workers presented a review on *in silico* studies focused on nuclear receptors, covering a range of approaches that included not only QSARs but also other methods such as molecular docking and dynamics, and considered the TR as the sole target related to the TH system [49]. In contrast, Vergauwen and co-workers presented a broader review focused on *in vivo*, *in vitro*, and *in silico* methods currently available for TH system disruption assessment [50]. However, their specific examination of *in silico* tools was confined to models available in open-source predictive tools (e.g., Danish (Q)SAR Database), leading to the identification of twelve models [50]. The present review addressed the current state-of-the-art of QSAR models published in the literature from 2010 and up to 2024 to predict potential TH system disruption by chemical substances. This allowed for a detailed characterisation of how this field has evolved over time, which type of TH system-related endpoints were modelled (and not) by these models, the main data sources used for model development, the modelling approaches, the applicability domain (AD) definitions, which types of chemicals have been assessed, which types of molecular descriptors have been selected as more relevant, and their mechanistic interpretations to suggest potential biological mechanisms. Mapping out the state-of-the-art on this topic is necessary to consolidate existing knowledge, identify research gaps, and offer a resource to guide future investigations. To provide the most up-to-date perspective on the topic, a separate paragraph is dedicated to key articles published between January and July 2025. The decision to treat these publications separately was made because 2025 is an incomplete year and a full comprehensive review of its literature would be premature.

5.2 Materials and methods

5.2.1 Criteria of inclusion and exclusion and literature collection

To meet the scope of this review, the following specific inclusion and exclusion criteria were predefined to collect relevant publications. (1) Original peer-reviewed research articles published from 2010 and up to 2024, where new QSAR models for predicting the potential TH system disruption by chemical substances were proposed. Original peer-reviewed research articles not proposing a new QSAR model (e.g., experimental studies, the application of unsupervised learning methods) were excluded. (2) Modelling efforts focused on predicting MIEs within AOPs for TH system disruption; the AOP network proposed by Noyes et al. was used as a reference [47]. MIEs, such as the induction of the constitutive androstane receptor (CAR), pregnane X receptor (PXR), aryl hydrocarbon receptor (AhR), and peroxisome proliferator-activated receptor (PPAR), were not considered in this review as they were not identified as being thyroid-specific by Dracheva et al. [51] in a study following that by Noyes et al. [47], and were also not addressed by the EURL ECVAM [33,48]. (3) From articles reporting multiple models for the same endpoint, only the QSARs explicitly identified as the best ones by the developers and/or applied for screening purposes within the same study were retained, thereby excluding QSARs arising from, e.g., different data partitioning, data imbalance handling techniques, and feature selection procedures (please note that a description about the effects of such approaches on models' performances was provided Section 5.3.5). (4) Original, peer-reviewed research articles focusing on QSAR development for a series of "selective ligands" in illness treatments or drug development were excluded. The same inclusion and exclusion criteria were applied to identify relevant articles published between January and July 2025.

The literature search was conducted using the Web of Science database, according to the inclusion and exclusion criteria. To obtain a more comprehensive collection of relevant publications, the literature search was conducted using both the full names and abbreviations of key biological targets (e.g., TTR, TPO) as keywords, rather than searching for each specific MIE (e.g., TTR binding, TPO inhibition) [47,48,51]. Hereafter in this review, each biological target will be referred to as the related MIE. The search strategy involved the following combinations of keywords: "thyroid system" AND "QSAR", "thyrotropin releasing hormone receptor" AND "QSAR", "TRHR" AND "QSAR", "thyroid stimulating hormone receptor" AND "QSAR",

"TSHR" AND "QSAR", "thyroperoxidase" AND "QSAR", "TPO" AND "QSAR", "sodium iodide symporter" AND "QSAR", "NIS" AND "QSAR", "type 1 deiodinase" AND "QSAR", "DIO1" AND "QSAR", "type 2 deiodinase" AND "QSAR", "DIO2" AND "QSAR", "type 3 deiodinase" AND "QSAR", "DIO3" AND "QSAR", "deiodinase" AND "QSAR", "DIO" AND "QSAR", "iodothyronine deiodinase" AND "QSAR", "IYD" AND "QSAR", "iodotyrosine deiodinase" AND "QSAR", "DUOX" AND "QSAR", "dual oxidase" AND "QSAR", "pendrin" AND "QSAR", "monocarboxylate transporter 8" AND "QSAR", "MCT8" AND "QSAR", "monocarboxylate transporter 10" AND "QSAR", "MCT10" AND "QSAR", "monocarboxylate transporter" AND "QSAR", "MCT" AND "QSAR", "organic anion transporter polypeptide 1C1" AND "QSAR", "OATP1C1" AND "QSAR", "organic anion transporter polypeptide 1A4" AND "QSAR", "OATP1A4" AND "QSAR", "organic anion transporter polypeptide" AND "QSAR", "OATP" AND "QSAR", "multidrug resistance protein 1" AND "QSAR", "MDR1" AND "QSAR", "multidrug resistance associated protein 2" AND "QSAR", "MRP2" AND "QSAR", "thyroid binding globulin" AND "QSAR", "TBG" AND "QSAR", "transthyretin" AND "QSAR", "TTR" AND "QSAR", "albumin" AND "QSAR", "thyroid receptor" AND "QSAR", "TR" AND "QSAR".

The selection of relevant publications from this search followed two main phases. An initial screening of titles and abstracts was conducted to assess relevance based on the inclusion and exclusion criteria. If relevance could not be determined from this step, a full-text analysis was performed.

5.3 Results and discussion

The final list comprised thirty publications including eighty-six distinct QSAR models. A summary is reported in Table 5.1, where studies are presented chronologically. Additional information is reported in Table S1 in the Supplementary Materials.

Table 5.1. Summary and main characteristics of selected QSARs. C: classification-based; R: regression-based; Primary: data generated as part of the same study; Secondary: data collected from the existing literature; ToxCast database: Toxicity Forecaster (ToxCast) database (https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast); Tox21 database: Toxicology in the 21st Century (Tox21) (https://tox21.gov/); Ref.: reference; n.s.: not specified.

Model ID	Ref.	Year	MIE	Algorithm	C or R	Chemical Class	Data Source Type	Data Source Literature Reference(s)
ID_1	[52]	2024	TBG	MLR	R	PBBs	Primary	[52]
ID_2	[52]	2024	TBG	MLR	R	PBBs and OH-PBBs	Primary	[52]
ID_3	[52]	2024	TBG	MLR	R	PBBs and 2OH-PBBs	Primary	[52]
ID_4	[52]	2024	TBG	MLR	R	PBBs, OH-PBBs, and 2OH-PBBs	Primary	[52]
ID_5	[53]	2024	TTR	MLR	R	Heterogeneous	Secondary	[54–60]
ID_6	[53]	2024	TTR	MLR	R	Heterogeneous	Secondary	[61–67]
ID_7	[53]	2024	TTR	MLR	R	Heterogeneous	Secondary	[68–89]
ID_8	[90]	2023	TR α	MLR	R	PFAS	Primary	[90]
ID_9	[90]	2023	TR β	MLR	R	PFAS	Primary	[90]
ID_10	[91]	2023	TR n.s.	LDA	C	OH-PCBs	Secondary	[92]
ID_11	[91]	2023	TR n.s.	LR	C	OH-PCBs	Secondary	[92]
ID_12	[93]	2023	Albumin	PLS	R	PFAS	Secondary	[94]
ID_13	[93]	2023	Albumin	LDA	C	PFAS	Secondary	[94]
ID_14	[93]	2023	Albumin	MLR	R	PFAS	Secondary	[94]
ID_15	[95]	2023	TSHR	RF	C	Heterogeneous	Tox21 database and secondary	[96–98]
ID_16	[51]	2022	TTR	RF	C	Heterogeneous	Secondary	[87]
ID_17	[51]	2022	TR β	RF	C	Heterogeneous	Tox21 database *	[99,100]
ID_18	[51]	2022	TR β	RF	C	Heterogeneous	Tox21 database *	[99,100]
ID_19	[51]	2022	TSHR	RF	C	Heterogeneous	Tox21 database *	[99,100]

ID_20	[51]	2022	TSHR	RF	C	Heterogeneous	Tox21 database *	[99,100]
ID_21	[51]	2022	TRHR	RF	C	Heterogeneous	Tox21 database *	[99,100]
ID_22	[51]	2022	DIO1	RF	C	Heterogeneous	ToxCast database **	[101]
ID_23	[51]	2022	DIO2	RF	C	Heterogeneous	ToxCast database **	[101]
ID_24	[51]	2022	DIO3	RF	C	Heterogeneous	ToxCast database **	[101]
ID_25	[51]	2022	NIS	RF	C	Heterogeneous	ToxCast database **	[102]
ID_26	[51]	2022	TPO	RF	C	Heterogeneous	ToxCast database **	[103]
ID_27	[104]	2022	TTR	RF	C	Heterogeneous	ChEMBL database ***	[105]
ID_28	[104]	2022	TR α	RF	C	Heterogeneous	ChEMBL database ***	[105]
ID_29	[104]	2022	TR β	RF	C	Heterogeneous	ChEMBL database ***	[105]
ID_30	[104]	2022	NIS	RF	C	Heterogeneous	ChEMBL database ***	[105]
ID_31	[106]	2022	TSHR	RF	C	Heterogeneous	Tox21 database	https://tripod.nih.gov/tox21/assays/
ID_32	[106]	2022	TSHR	RF	C	Heterogeneous	Tox21 database	https://tripod.nih.gov/tox21/assays/
ID_33	[106]	2022	TSHR	XGB	C	Heterogeneous	Tox21 database	https://tripod.nih.gov/tox21/assays/
ID_34	[106]	2022	TSHR	LR	C	Heterogeneous	Tox21 database	https://tripod.nih.gov/tox21/assays/
ID_35	[106]	2022	TSHR	XGB	R	Heterogeneous	Tox21 database	https://tripod.nih.gov/tox21/assays/
ID_36	[107]	2022	TPO	XGB	C	Heterogeneous	ToxCast database and secondary **	[103,108–111]
ID_37	[107]	2022	TPO	Hard Voting	C	Heterogeneous	ToxCast database and secondary **	[103,108–111]
ID_38	[107]	2022	TPO	Soft Voting	C	Heterogeneous	ToxCast database and secondary **	[103,108–111]

ID_39	[112]	2022	TR β	MLR	R	PCNs	Primary	[112]
ID_40	[113]	2023	TR β	RF	C	Heterogeneous	Tox21 database	National Center for Biotechnology Information. PubChem Database. Source = 824, AID = 743067, https://pubchem.ncbi.nlm.nih.gov/bioassay/743067 (accessed 13 May 2021)
ID_41	[59]	2021	TTR	MLR	R	Halogenated phenols and thiophenols	Primary and Secondary	[57,59]
ID_42	[114]	2021	TTR	kNN	C	Heterogeneous	Secondary	[54–59,61,62,64,68,70–73,75,78–89,115–118]
ID_43	[114]	2021	TTR	kNN	C	Heterogeneous	Secondary	[54–59,61,62,64,68,70–73,75,78–89,115–118]
ID_44	[114]	2021	TTR	kNN	C	Heterogeneous	Secondary	[54–59,61,62,64,68,70–73,75,78–89,115–118]
ID_45	[114]	2021	TTR	kNN	C	Heterogeneous	Secondary	[54–59,61,62,64,68,70–73,75,78–89,115–118]
ID_46	[114]	2021	TTR	kNN	C	Heterogeneous	Secondary	[54–59,61,62,64,68,70–73,75,78–89,115–118]
ID_47	[114]	2021	TTR	MLR	R	Heterogeneous	Secondary	[61,70–73,75,78–88]
ID_48	[114]	2021	TTR	MLR	R	Heterogeneous	Secondary	[57–59]
ID_49	[114]	2021	TTR	kNN	R	Heterogeneous	Secondary	[61,70–73,75,78–88]
ID_50	[114]	2021	TTR	kNN	R	Heterogeneous	Secondary	[57–59]
ID_51	[119]	2021	TR n.s.	RF	C	Heterogeneous	ToxCast database	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC.
ID_52	[119]	2021	TSHR	RF	C	Heterogeneous	ToxCast database	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC.
ID_53	[119]	2021	TSHR	NN	C	Heterogeneous	ToxCast database	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC.

ID_54	[119]	2021	TPO	XGB	C	Heterogeneous	ToxCast database **	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC. and [103]
ID_55	[119]	2021	TRHR	SVM	C	Heterogeneous	ToxCast database	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC.
ID_56	[119]	2021	DIO1	SVM	C	Heterogeneous	ToxCast database **	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC. and [101]
ID_57	[119]	2021	DIO2	SVM	C	Heterogeneous	ToxCast database **	[101]
ID_58	[119]	2021	DIO3	NN	C	Heterogeneous	ToxCast database **	[101]
ID_59	[119]	2021	NIS	LR	C	Heterogeneous	ToxCast database **	Cited as ToxCast and Tox21 Summary Files for invitroDBv3.2, U.S. EPA, Washington, DC. and [102]
ID_60	[120]	2021	TPO	kNN	C	Heterogeneous	ToxCast database **	[103,121]
ID_61	[120]	2021	TPO	RF	C	Heterogeneous	ToxCast database **	[103,121]
ID_62	[57]	2019	TTR	MLR	R	Phenolic DBPs	Primary	[57]
ID_63	[122]	2018	TR β	SVM	C	PCBs	Primary	[122]
ID_64	[122]	2018	TR β	LDA	C	PCBs	Primary	[122]
ID_65	[123]	2018	TR n.s.	SVM	C	PCBs and PBDEs	Secondary	[124–132]
ID_66	[133]	2017	TTR	LDA	C	PFCs	Secondary	[82]
ID_67	[133]	2017	TTR	MLR	R	PFCs	Secondary	[82]
ID_68	[121]	2017	TPO	PLR	C	Heterogeneous	ToxCast database **	[103,134–136]
ID_69	[121]	2017	TPO	PLR	C	Heterogeneous	ToxCast database **	[103,134–136]
ID_70	[87]	2015	TTR	kNN	C	Heterogeneous	Secondary	[88]
ID_71	[137]	2015	TTR	ASNN	C	Heterogeneous	Secondary	[88]
ID_72	[138]	2015	TR β	Monte Carlo	R	Heterogeneous	Secondary	[139]
ID_73	[138]	2015	TR β	Monte Carlo	R	Heterogeneous	Secondary	[139]
ID_74	[138]	2015	TR β	Monte Carlo	R	Heterogeneous	Secondary	[139]

ID_75	[139]	2014	TR β	RF	R	Heterogeneous	ChEMBL database ***	[140]
ID_76	[139]	2014	TR β	RF	R	Heterogeneous	Secondary	[141–143]
ID_77	[139]	2014	TR β	RF	C	Heterogeneous	ChEMBL database ***	[140]
ID_78	[144]	2013	TTR	kNN	C	PFCs and BFRs	Secondary	[78,80,82]
ID_79	[144]	2013	TTR	MLR	R	PFCs and BFRs	Secondary	[78,80,82]
ID_80	[145]	2012	TTR	kNN	C	PFCs	Secondary	[82]
ID_81	[145]	2012	TTR	kNN	C	PFCs	Secondary	[82]
ID_82	[145]	2012	TTR	kNN	C	PFCs	Secondary	[82]
ID_83	[145]	2012	TTR	kNN	C	PFCs	Secondary	[82]
ID_84	[146]	2011	TTR	kNN	C	BFRs	Secondary	[78,80]
ID_85	[147]	2010	TTR	MLR	R	BFRs	Secondary	[78,80]
ID_86	[148]	2010	TR β	PLS	R	OH-PBDEs	Primary	[148]

* Tox21 served as a data source but it was cited as [99,100]. ** Although cited as [101–103] or [134–136], this review will refer to them as the ToxCast data source as described in the referenced papers. *** ChEMBL served as a data source but it was cited as [105] or [140].

5.3.1 Temporal trends

Figure 5.1 illustrates the number and distribution of the selected QSAR models and papers over time. Despite minor fluctuations, modelling efforts remained relatively stable since 2010 until 2020, followed by a noticeable surge in the period 2021–2022 and by a slight decrease in 2023–2024. Sixteen out of the thirty papers selected in this review were published in the period 2021–2024, suggesting a recent acceleration of research into TH system disruption using QSAR-based approaches. Before this, the field was characterised by notably sparser publications, with fourteen papers appearing over the ten years from 2010 to 2020. Notably, no relevant publications were detected in 2016 and 2020, which could signify periods of reduced research focus (e.g., the impact of the COVID-19 pandemic), or a shift in research priorities. The number of developed QSAR models mirrors this trend. Indeed, while a year-to-year fluctuation was observed up to 2020, over 70% of the total QSARs were published within the last four years, with a pronounced surge occurring in 2021 and 2022. The number of QSAR models exceeding the number of publications is largely attributed to the increasing practice of proposing multiple models within a single publication, often addressing, for instance, different endpoints, descriptor types, and/or methodological approaches. These findings could be mainly attributed to the growing availability of publicly available high-throughput screening (HTS) data for multiple thyroid-related endpoints, such as those from large-scale projects like Toxicity Forecaster (ToxCast)

(https://www.epa.gov/comptox-tools/toxicity-forecasting-toxcast) and Toxicology in the 21st Century (Tox21) (https://tox21.gov/).

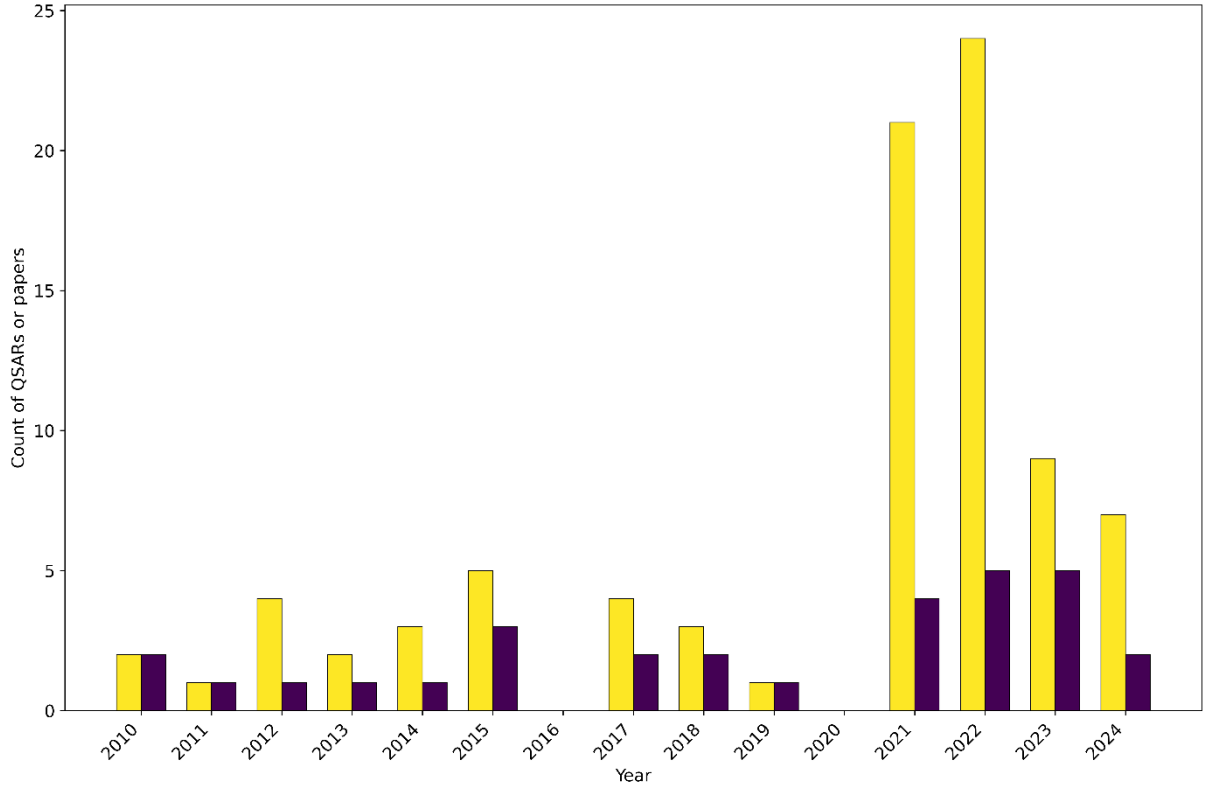

Figure 5.1. Annual distribution of QSAR models (yellow bars) and papers (purple bars).

5.3.2 Modelled MIEs

The selected QSAR models were developed for eleven different MIEs for TH system disruption, which represent only a subset of the over twenty described by Noyes et al. [47]. MIEs regarding DUOX, IYD, and pendrin inhibition, as well as those related to cellular TH transport (i.e., MCT8, MCT10, OATP1C1, OATP1A4, MDR1, and MRP2), have never been addressed by QSAR modelling.

As illustrated in Figure 5.2, a predominant focus was placed on TR and TTR, which together account for 57% of all the QSARs included in this review. This large number could be attributed to the widespread availability of *in vitro* data for these MIEs and to their established mechanistic links with TH system disruption [7]. While less frequently modelled than TTR and TR, targets like TSHR and TPO were still relatively well represented. The modelling of TPO, which is a key enzyme for THs synthesis, and TSHR, which is a protein that regulates thyroid gland function, highlighted an

expanding scope of investigation beyond just TH distribution or nuclear receptor binding reflected by, respectively, TTR and TR. In contrast, other important MIEs remained significantly poorly addressed, highlighting the notable gaps in the current research in the field. The critical roles of albumin, TBG, NIS, TRHR, and the three deiodinases (DIO 1, 2, and 3) in TH synthesis, distribution, and metabolism are well established [31,47]. However, despite their recognised relevance, the scarcity of QSAR research for these targets pointed out potential challenges, such as poor data availability or a limited interest or awareness among QSAR developers. This almost-negligible modelling effort for these MIEs indicates a significant opportunity for future research and QSAR model development.

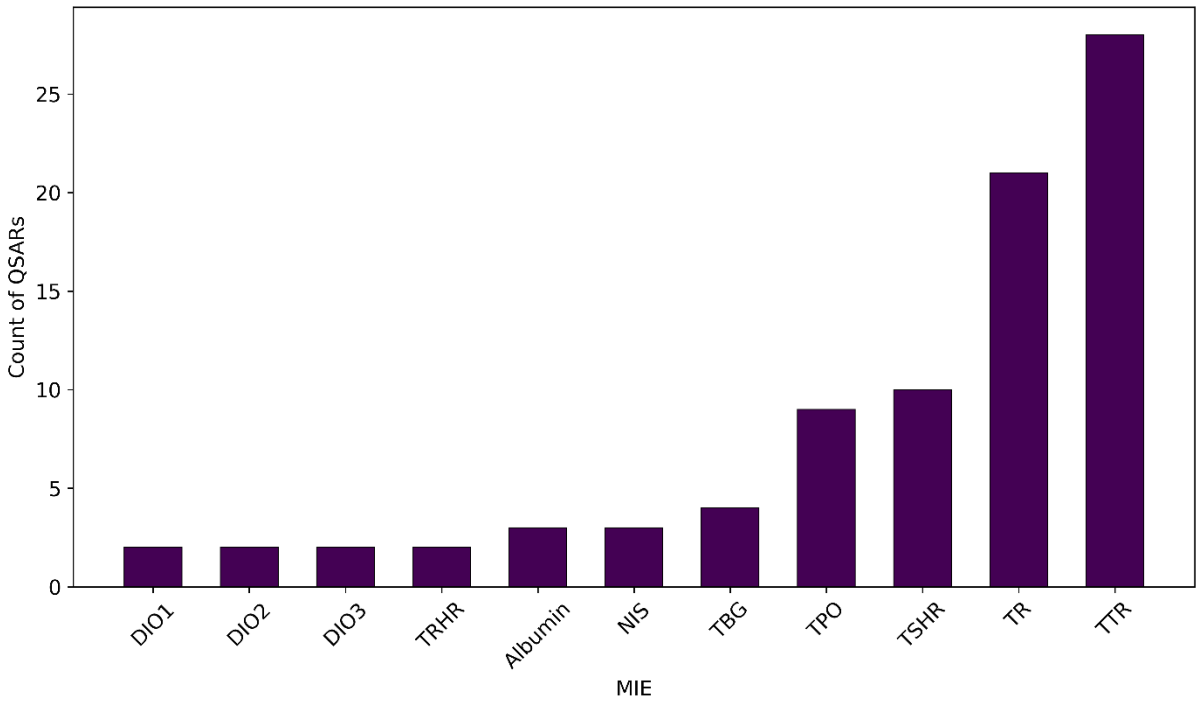

Figure 5.2. Count of QSAR models developed for each MIE.

As illustrated in Figure 5.3, TR and TTR were consistently modelled throughout the entire study period, reflecting their long-standing recognition as key targets for TH system disruption assessment. A shift in research focus is evident from 2021 onward, with a significant diversification of modelled MIEs. Specifically, the modelling efforts on TSHR, TPO, NIS, TRHR, and deiodinases (DIO1, DIO2, DIO3), though less numerous overall, were distinctly concentrated in 2021 and 2022. This concentrated activity, however, largely stemmed from two studies by Dracheva et al. [51] and de

Lomana et al. [119], where multiple endpoints were addressed in the same publication. QSARs addressing other important TH distributor proteins, i.e., TBG and albumin, were only published in the last two years.

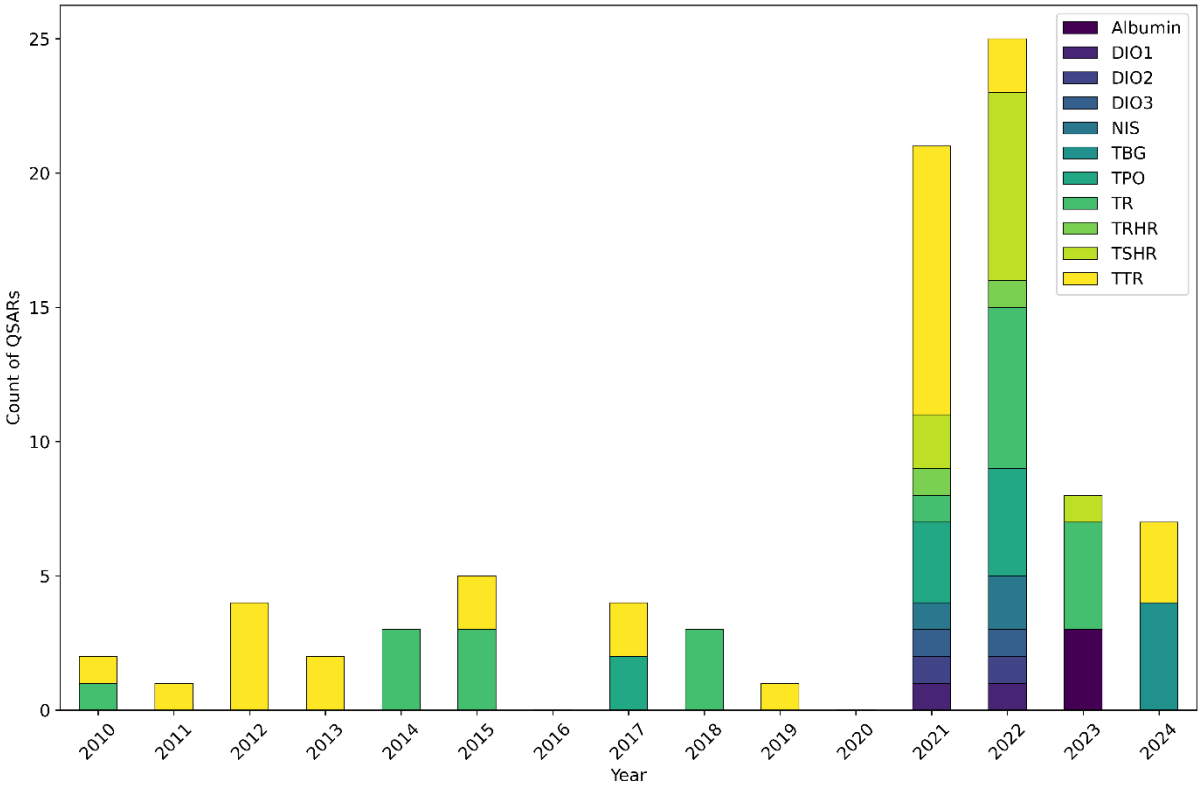

Figure 5.3. Annual distribution of QSAR models, categorised by MIE.

As discussed in Section 5.3.1, this trend of diversification and the surge in the 2021–2022 biennium are likely linked to the growing availability and accessibility of HTS data. Prior to 2021, the scarcity of QSAR studies for MIEs other than TR and TTR likely stemmed from a combination of factors: a scarcity of available experimental data (for instance, Gadaleta et al. [104] pointed out that MIEs such as MCT8, MCT10, and OATP1C1 lacked sufficient active compounds in the ChEMBL database to be used for modelling purposes) and a complexity of developing suitable and validated assays for their generation, or a lower awareness of the mechanistic role of these MIEs in TH system disruption. It is worth highlighting that no *in vitro* assays for TH system disruption have yet been validated by the OECD [33,48,149,150], which might slow data generation. The relatively recent publication of the AOP network for TH system disruption by Noyes and colleagues [47] likely played a crucial role. By providing a more structured understanding of these diverse pathways, it stimulated research into

previously underexplored MIEs. The growing number of publications and QSARs covering multiple MIEs underscored the increasing awareness of the multifaceted and interconnected nature of TH system disruption.

5.3.3 Data sources

As detailed in Table 5.1, the QSARs selected for this review were based on data from three main source types: (1) primary sources, where data was generated as part of the same study; (2) secondary sources, where data was collected from the existing literature; (3) publicly available databases (i.e., ToxCast, Tox21, and ChEMBL). In most cases, these sources were used individually, while in others, they were combined (Figure 5.4).

The data source reference(s) used to develop each QSAR are reported in Table 5.1. The data included in publicly available databases served as unique data sources for developing thirty-five distinct QSARs, representing approximately 41% of the total. MIEs covered by these QSARs included TTR, TR, TSHR, TPO, TRHR, NIS, and the three deiodinases. With a single exception [121], all of the studies using data from the ToxCast and Tox21 projects were published from 2021 to 2023, proposing all of the available QSARs addressing TPO, NIS, TSHR, TRHR, DIO1, DIO2, and DIO3. As previously discussed, these findings were linked to the growing availability and accessibility of comprehensive HTS datasets. This data availability, combined with an increasing awareness of the critical roles these targets play within the TH system, has broadened the scope of QSAR investigations for TH system disruption assessment.

In contrast, primary and secondary data sources were used alone for the development of forty-six (53%) QSARs. The consistent use of primary and secondary data sources from the literature throughout the entire study period underscored their sustained importance. These models covered a stricter range of MIEs, such as TR and TH distributor proteins TTR, TBG, and albumin, underscoring limited data availability or utilisation for other MIEs. Whilst the majority of these QSARs were developed using data from *in vitro* experiments, Kowalska et al. [90] and Yang et al. [52] developed a total of six QSARs to predict binding energies to TTR and TBG, respectively. Binding energies used for models' development were generated within the same studies through molecular docking and dynamic simulations and used as dependent variables. The successful application of integrated *in silico* approaches highlighted their utility as an effective strategy when experimental data from *in vivo* or *in vitro*

studies is limited or entirely lacking, further enabling the exploration of complex molecular interactions that might be otherwise inaccessible.

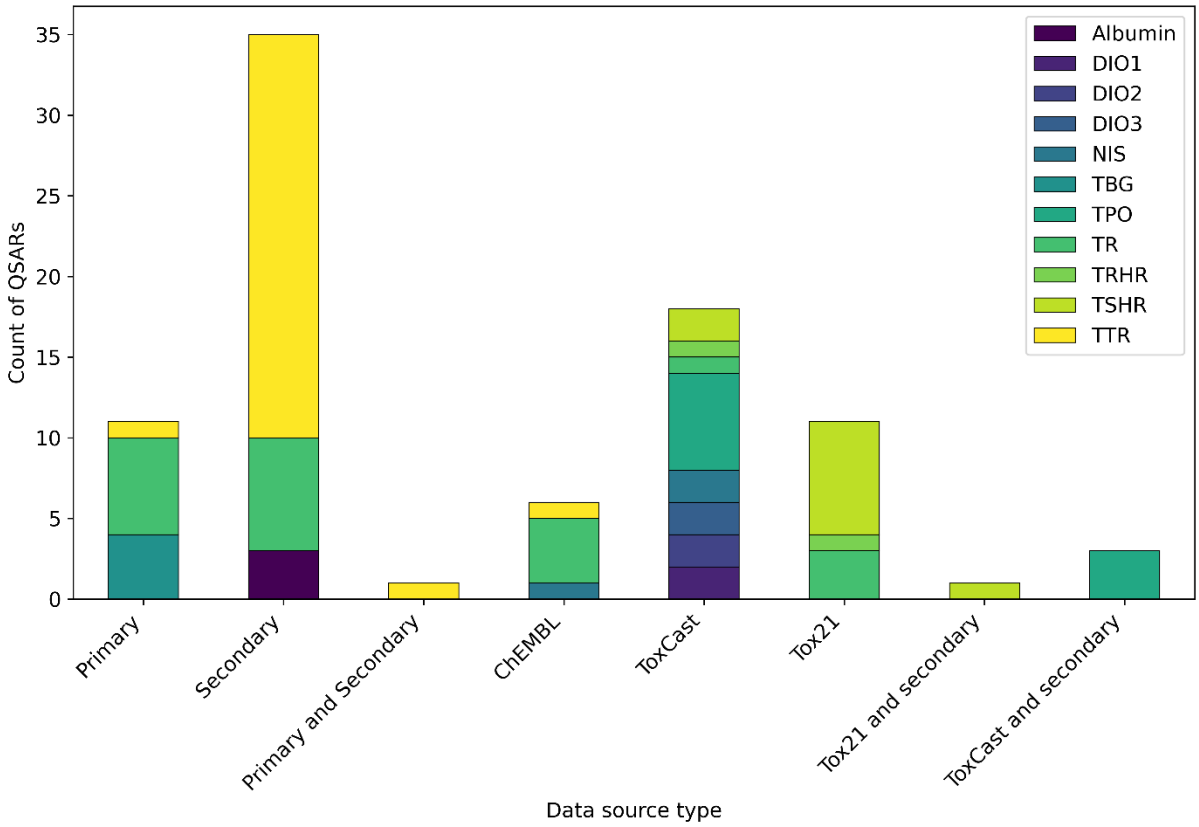

Figure 5.4. Count of QSAR models based on data source type, categorised by MIE.

A key aspect across the studies was data transparency. The data sources and data used for model development were consistently made available, either directly within the publications or through adequately referenced sources. This commitment to data availability aligned with the FAIR (findable, accessible, interoperable, reusable) principles for data sharing [151], thereby optimising data reuse for future research.

5.3.4 Chemical classes

The datasets used for QSARs training and validation included either structurally heterogenous chemicals or class-specific chemicals.

Structurally heterogeneous datasets were used for approximately 67% of the QSARs. These datasets primarily consisted of organic chemicals, encompassing a mix of environmental pollutants, natural compounds, and, occasionally, drugs. The sizes of such datasets varied considerably, from 41 to 8682 compounds. About 83% of these QSARs were published within the last four years, reflecting the spreading availability of HTS data, as previously described. As illustrated in Figure 5.5, all the endpoints were addressed using heterogeneous datasets, with the exception of TBG and albumin.

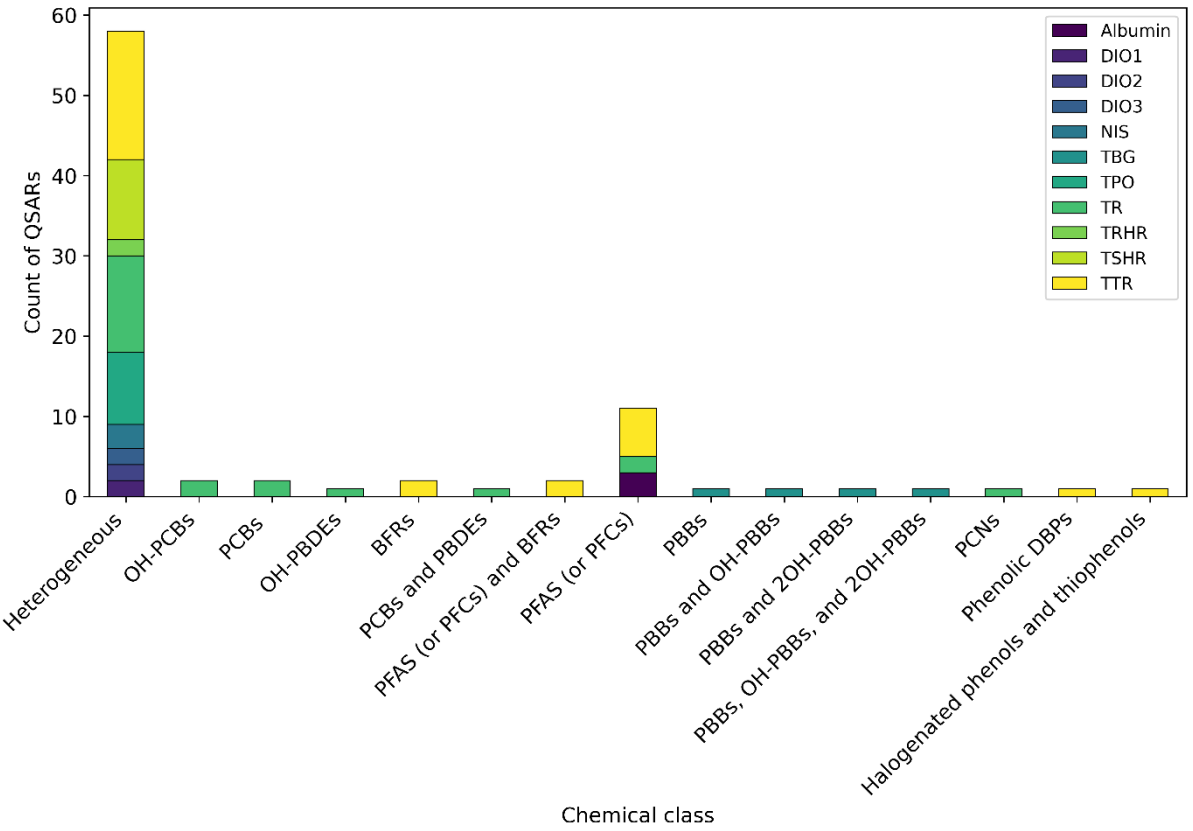

Figure 5.5. Count of QSAR models based on structurally heterogeneous or class-specific datasets, categorised by MIE.

In contrast, only a limited number of chemical classes have been tested and modelled for TH system disruption, addressing a limited number of MIEs. These datasets

primarily focused on environmental pollutants of known concern, including PCBs and their hydroxylated metabolites, PBDEs and their hydroxylated metabolites, PCNs, halogenated phenols and thiophenols, phenolic DBPs, PFAS (often referred to as PFCs), and PBBs and their hydroxylated metabolites. The sizes of these datasets were generally smaller compared with the structurally heterogeneous ones, ranging from 17 to 107 compounds. Furthermore, these data were exclusively generated within the same study or retrieved from the existing literature, hence were never extracted from databases. Notably, only TR and TH distributor proteins (i.e., TTR, TBG, and albumin) were modelled using these datasets, underscoring limited data availability or the utilisation of specific class data for other thyroid-related endpoints. It is also important to highlight that almost half of these QSARs were published within the last four years. This trend suggested that, despite the increasing availability of HTS data, the reliance on data published in the literature by independent research groups remained critically important.

Although certain compounds, like bisphenol derivatives, phthalates, various pesticides, and constituents of personal care products have been experimentally identified as THSDCs [22,25,26,152], many others within these same classes remain poorly addressed. This lack of data is concerning because structural similarity among compounds within the same class may suggest a similar toxic potential. This highlighted a strong need for additional *in silico* or *in vitro* efforts to generate more data for these and other chemical categories for specific MIEs. Broadening the chemical space coverage for each of these chemical categories would be essential to develop new, specific QSAR models, enabling a more robust hazard assessment for entire groups of compounds. Building on the successful application of integrated *in silico* approaches by Kowalska et al. [90] and Yang et al. [52], as described in Section 5.3.3, a similar approach could be an effective strategy to address other MIEs for specific chemical classes.

Generally, the use of heterogeneous datasets can improve a model's AD coverage and generalizability for large screening applications. In contrast, local QSARs, which are specifically designed for specific classes of compounds, are often preferred for their ability to more accurately capture subtle structural differences and specific structure–activity relationships. This can lead to (potentially) more reliable predictions within that defined chemical space. Therefore, the choice between using global or local QSAR models depends on the specific application purposes. Furthermore, the inherent complexity of heterogeneous data can hinder the mechanistic interpretation of

molecular descriptors (see Section 5.3.8). When a model is trained on a wide array of chemical structures, it is more challenging to pinpoint the exact structural features or physicochemical properties responsible for a particular activity. This is in contrast to datasets of specific classes, where a clearer structure–activity relationship can emerge, making interpretation more straightforward.

5.3.5 Modelling approaches

A wide variety of modelling algorithms are available for QSAR model development. These range from traditional methodologies, such as MLR and LDA, to more complex machine learning methodologies, such as NN and SVM [46,153]. The choice of algorithm generally depends on the complexity of the data and the desired interpretability of the model. Thus, the landscape of algorithms for QSAR development lacks a universally accepted solution, as each method presents its own set of strengths and limitations.

As illustrated in Figure 5.6, different modelling algorithms and approaches were identified across the papers. Over two-thirds of the QSARs selected in this review (67%) were designed for classification, a preference largely driven by the nature of HTS data. Large-scale projects like ToxCast and Tox21 generate vast datasets, where the effect on a biological target by compounds is often reported with a simple categorical outcome, i.e., "active" or "inactive". This format has consequently led to a shift in QSAR modelling for TH system disruption, favouring classification-based approaches over regression-based ones.

RF was the most frequently used algorithm, followed by MLR and kNN. Overall, RF, kNN, and MLR were used to develop a total of fifty-eight different QSARs, corresponding to approximately 67% of the total models. It is important to highlight that a single study by Dracheva et al. [51] utilised RF to develop eleven different QSARs for the prediction of nine MIEs, which significantly influenced the overall count of RF applications.

The majority of studies concentrated on a single, well-defined modelling strategy, while a few explored more comprehensive approaches, systematically exploring combinations of algorithms, descriptor types, or class-balancing techniques to achieve the best possible performance. While a comprehensive comparative analysis of the predictive models' performances would be highly valuable, it fell outside the scope of this review, as it was hindered by the following two key reasons. Firstly, the

distribution of available models was highly imbalanced. While MIEs like TTR and TR have been extensively studied with multiple QSARs, others have been addressed by a few, or even no, models. Secondly, cross-study comparisons of models' performances can be performed only when the same dataset and data processing technique are used [154], meaning that simply looking at the statistical metrics of QSARs from different papers would be inappropriate to determine which modelling approach is truly superior. For example, Schür et al. recently reviewed predictive ecotoxicology studies and concluded that no existing studies were truly comparable due to inconsistent methodologies regarding datasets, data processing, and performance statistical metrics [155]. This finding could also be applicable to the broader toxicological context. Therefore, the focus of this section was placed on studies that directly explored various modelling approaches, in terms of algorithms, descriptor types, or data-balancing methods, on a single, consistent dataset. This approach allowed the authors to conduct a reliable assessment of which specific methodology yielded the best predictive results.

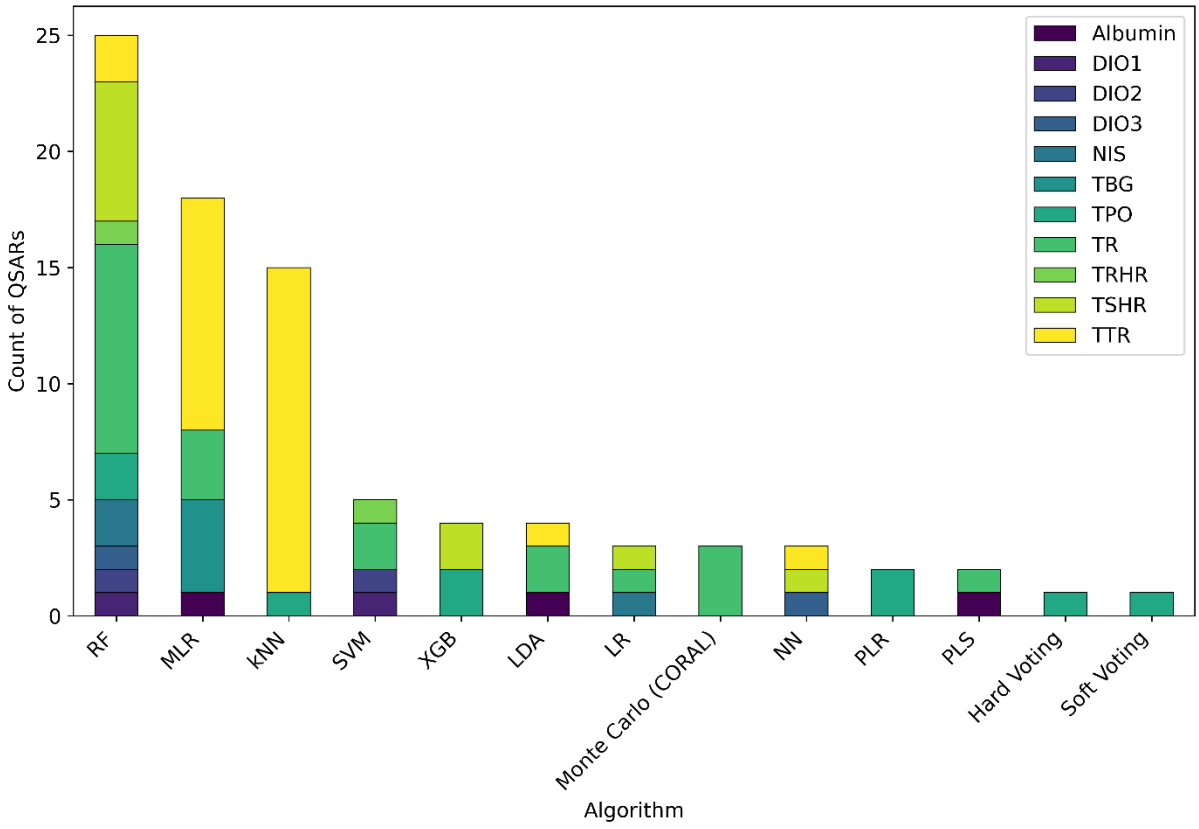

Figure 5.6. Count of QSAR models based on the modelling algorithm, categorised by MIE.

All of the models for TPO inhibition were developed using structurally heterogenous datasets of chemicals (see Table 5.1 and Figure 5.5). Rosenberg et al. [121] developed

two robust QSAR models using PLR, named QSAR1 and QSAR2, using an initial selection of predefined molecular descriptors and training set-dependent scaffolds. The authors evaluated seven different modelling strategies, including approaches that used scaffolds and those that did not, in both single and composite models. The most successful strategy was a composite model that uniquely combined a single, unbalanced model with balanced sub-models from a composite one. This strategy was found to be particularly effective in handling the challenges posed by imbalanced datasets, and led to the final QSAR1 and QSAR2 models (with a cross-validation balanced accuracy equal to 80.6% and 82.7%, respectively). Similarly, Seo et al. [107] developed binary, ternary, and quaternary QSAR models. They applied multiple algorithms, such as RF, SVM, artificial NN, Adaptive Boosting (AdaB) and XGB, and hard- and soft-voting classifiers. Each algorithm was combined with multiple categories of fingerprints (FPs) (e.g., Morgan FPs, Atom Pair Count FPs) and dimensionality reduction techniques (i.e., principal component analysis (PCA) and LDA) to address overfitting. The Atom Pair Count FPs was the best-performing FP, whereas the best-performing models in the binary, ternary, and quaternary models were the hard-voting classifier, XGB with LDA, and soft-voting classifier, respectively (test scores equal to 0.66, 0.51, and 0.52, respectively). Gadaleta et al. [120] applied multiple algorithms, including SVM, balanced RF, RF, and kNN, and explored different partitioning schemes to stratify and select active compounds in different ways. The top-performing models were based on balanced RF and kNN using a dataset that excluded compounds with an ambiguous active categorization. The models achieved a balanced accuracy of 76–78% on external data, which resulted as a performance comparable to the reported experimental variability of the assay used to generate modelled data.

Regarding TR binding, Bai et al. [122] developed classification QSARs based on twenty-two PCBs using LDA and SVM. Both showed strong and equal accuracy in the training set (88.2%), with the SVM model exhibiting a greater accuracy in the test set, equal to 80%. Akinola et al. [91] developed classification models applying LR and LDA on a dataset of sixty-eight OH-PCBs, showing that both methods performed identically (accuracies in the training set and in the test set equal to 84.3% and 76.5%, respectively). Yan et al. [123] developed ternary classification models applying LDA, classification and regression trees (CART), and SVM on a dataset of structurally heterogenous compounds. SVM proved to be the optimal algorithm, with a total accuracy in the training and test set equal to 81.4% and 76.5%, respectively. Sapounidou et al. [113] proposed a comprehensive set of twenty-three QSAR models for various MIEs related

to endocrine disruption, including TRβ binding, utilising the conformal prediction (CP) framework combined with RF as the modelling algorithm. Five different data-balancing techniques were employed (for more details, see below), with CP providing the best one. A balanced accuracy equal to 0.78 was achieved.

As for TPO, all of the models for TSHR inhibition were developed using datasets of structurally heterogenous chemicals. Xu and colleagues [106] developed binary classification models comparing three different algorithms: RF, XGB, and LR. Both RF and XGB models showed good predictive performances, with balanced accuracies of 0.85 and 0.84, respectively. The authors further developed a simplified RF model using the seven most influential descriptors, which maintained strong performance (balanced accuracy equal to 0.83). Additionally, they first developed a regression model using MLR, which yielded an R2 of 0.35. Therefore, a regression model using XGB was developed and the R2 increased up to 0.65. Later, Liu et al. [95] explored various combinations of seven molecular representations (including different types of FPs and Mordred descriptors) and four algorithms (RF, SVM, multilayer perceptron, and graph attention network). The best-performing model was a RF using PubChem FPs, which achieved a balanced accuracy of 0.94 on the validation set.

Regarding TTR binding, Zhang et al. [87] developed QSAR classification models applying kNN, PLS discriminant analysis (PLS-DA), and SVM. The kNN model, with a k-value of 4, showed the best performance, achieving the highest correct classification rate during both internal and external validation (0.88 and 0.82, respectively). Similarly, Rybacka et al. [137] tested seven different machine learning methods (MLR, PLS, associative NN, kNN, RF, SVM, and fast stepwise (stagewise) multivariate linear regression) and five distinct descriptor sets. The best result was obtained by combining the associative NN algorithm with Dragon descriptors, which achieved a balanced accuracy equal to 89%.

Finally, de Lomana et al. [119] used five different algorithms (i.e., LR, RF, XGB, SVM, and NN) in combination with three class-balancing techniques (see below for more details) to predict multiple MIEs. All algorithms performed similarly, with a tendency for the models trained on over-sampled data to achieve better results. Balanced accuracies ranged from 0.68 to 0.82 for different endpoints.

Although the use of diverse algorithms was evident (Figure 5.6), no clear temporal trend was observed in the type of modelling algorithms employed. This suggested a consistent application of both established and newer algorithms across different publication years, rather than a gradual shift toward more complex techniques.

Interestingly, despite the recent advancements in machine learning and deep learning approaches, classical algorithms such as MLR, LDA, and PLS continued to be widely used, given their interpretability and simplicity. Indeed, their added value stems from providing easily understandable models that offer direct insights into the structural features driving the activity, which is a key aspect to enhance confidence with QSARs. On the contrary, the "black box" nature of more complex algorithms makes them less transparent and, if not adequately controlled, potentially more susceptible to overfitting [156]. Therefore, an important research direction is to leverage the power of complex algorithms by focusing on developing methods that enhance their interpretability and transparency, thereby increasing user confidence and facilitating their broader adoption.

Establishing a clear link between the algorithm type and a specific endpoint proved challenging, as most endpoints have been assessed by a few, or even no, QSARs. Regarding TTR and TR, the two most modelled endpoints, algorithms capable of handling linear relationships (e.g., MLR) and non-linear relationships (e.g., kNN) between independent and dependent variables were both utilised, with a slight preference for the second group.

An additional methodological aspect observed across the studies was the application of class-balancing strategies. This is crucial to address class imbalance, where one class (e.g., inactive compounds) is much more common than another (e.g., active compounds) in a training dataset. This imbalance is frequently found in data from databases or generated through HTS and can cause a model to become biased toward the majority class, leading to a poor performance with the minority class. This is especially critical in hazard prediction, where mistakenly predicting a dangerous compound as safe is a far more serious error than the opposite. Several effective strategies exist and were observed in the reviewed studies. Sapounidou et al. [113] in combination with RF, explored five different data-balancing techniques: CP, equal size sampling (under-sampling), over-sampling by duplication, synthetic minority over-sampling technique, and random over-sampling examples. As described above, the use of CP was the best choice. de Lomana et al. [119] combined five different algorithms (i.e., LR, RF, XGB, SVM, and NN) and three class-balancing techniques: weight balancing, over-sampling, and under-sampling. The models trained on over-sampled data achieved better results. Xu et al. [106] employed the synthetic minority over-sampling technique-edited nearest neighbours (SMOTEENN) technique, which combines over-sampling the minority class samples with under-sampling the majority

class samples to achieve a more balanced distribution. Gadaleta et al. [104,120] developed models using balanced RF, which is an adaptation of the more traditional RF that incorporates the internal balancing of categories. Finally, Liu et al. [95] employed a threshold moving method. The increasing volume of HTS data highlights the critical importance of effective class-balancing strategies for enhancing the robustness and reliability of models built on these datasets.

5.3.6 Validation strategies

Validation stands as a crucial step in QSAR model development, ensuring the appropriateness of goodness-of-fit, overall robustness, and predictive ability, thereby ultimately maximising the model's reliability [157,158,159]. Validation procedures can be distinguished as internal and external. Internal validation is conducted to evaluate the robustness and the predictive ability of a QSAR on the data from which it was developed (i.e., training set). External validation, on the other hand, is conducted to evaluate the actual predictive ability of a QSAR on data not used for its development (i.e., test set). Thus, external validation is of key importance as it assesses the model's true predictive power using unseen data [157,158,160]. Although the best strategy to perform external validation involves the use of completely new and independent datasets, obtaining these datasets is often challenging given the scarcity of available experimental data. Therefore, a common practice is to partition the available data into a training set and into a test set (i.e., dataset splitting) [157,158,161].

All of the QSARs reviewed in this study underwent some form of internal and/or external validation. Internal validation was performed to evaluate the robustness of seventy-two QSARs, accounting for approximately 84% of the total. The most frequent internal validation strategy was the *k*-fold cross-validation (CV), which was used in fifty-five instances. This approach involves splitting the training set into *k* equally sized groups. The method iteratively trains a model on $k - 1$ groups and validates it on the remaining group. This is repeated *k* times, such that each group serves as the validation set once [162]. In this review, *k* values were typically set to 2, 5, or 10. In some instances, this strategy was often referred to as leave-more-out CV (LMO CV) and as leave-one-out CV (LOO CV). The latter is the simplest case of *k*-fold CV, where each compound of the training set is removed one at time, and it was employed in twenty instances. Finally, the stratified bagging method was used in one instance [137], where *k*-fold CV was also tested. Additionally, regression-based QSARs often underwent further internal validation strategies, such as the QUIK rule [163] to detect high predictor

collinearity, Y-randomisation to detect chance correlations [164], and the use of the bootstrapping coefficient [165]. Details about each model are reported in Table S1. External validation was performed to evaluate the predictive ability of eighty-three QSARs, representing almost all of them. It is important to highlight that for the three QSARs where external validation was not conducted, this omission was not an oversight, instead it was intentionally not performed and adequately justified by the authors [93,121]. For example, Rosenberg et al. [121] proposed two QSARs for TPO inhibition, named QSAR1 and QSAR2, which were developed using two independent datasets. QSAR1 was developed using one dataset as the training set and the other one as the test set for external validation. Instead, QSAR2 was developed by merging both datasets to form a larger training set: whilst QSAR2 was developed using the same modelling method and CV approaches as QSAR1, it purposely lacked external validation. Both QSARs showed good performances and were applied to broader screening purposes. In the study by Gallagher et al. [93], given the small size of the dataset (twenty-two compounds), the "Small Dataset Modeler" tool proposed by Ambure et al. [166] was utilised to facilitate an exhaustive double CV approach that uses the entire dataset without requiring splitting it into a training set and a test set, making it an effective and suitable solution to validate QSARs based on limited data. Beyond these three specific exceptions, data partitioning into a training and test set was performed by employing various splitting strategies. Random splitting is a frequently adopted strategy [167], and its prevalence was also observed in this review, where it was applied in sixty-three instances. While straightforward, this procedure can lead to an uneven data distribution particularly when dealing with small-sized datasets or with skewed class distributions [168,169,170]. This imbalance might ultimately result in training and test sets that deviate from the representativeness suggested by Golbraikh et al. [171], who argued that using rationally selected training and test set can enhance QSAR reliability [171]. Alternative partitioning strategies have been proposed and used for a strategic selection of training and test set compounds [157,168]. In this review, strategies alternative to random splitting included those based on sorted response variables [53,90,133,144] or on the Kennard–Stone algorithm [87,91,172]. The details about each model are reported in Table S1.

The diverse array of splitting strategies reflected the fact that there is not a single and widely considered ideal partitioning scheme. Instead, the choice depends on the specific dataset type, its size, and the modelling methodology employed in the study [168]. Encouragingly, with only a few noted exceptions, the predominant practice across the reviewed studies involved the combined application of both internal and

external validation. This robust approach, which was utilised in 80% of the QSAR models, suggested a strong commitment within the field to ensure their validity and reliability.

5.3.7 Applicability domain

A single QSAR model cannot accurately predict the entire chemical universe [173]. Thus, each QSAR needs to be associated with a clearly defined AD. This domain determines whether a QSAR model can provide reliable or unreliable predictions (i.e., extrapolations) based on the structural, physicochemical, and response information present in the training set of the model [157,158,174]. No single, universally accepted method exists for defining the AD of a QSAR model. Instead, a range of methodologies are utilised, each offering a distinct approach [175,176]. These methods can differ in their restrictiveness and can yield either categorical outcomes (e.g., a simple "in" or "out" of the AD) or continuous values (e.g., distance) quantifying the relative position of a compound to the AD boundaries or centre [157].

An alarming finding was that the AD was not explicitly defined for thirty-two QSARs, accounting for approximately 37% of the total models. Among these, it is worth highlighting the studies by Bai et al. [122] and by Akinola et al. [91]. Bai et al. [122] developed two QSARs using a training set of twenty-two PCBs, which were then applied to predict TR binding for the remaining PCBs congeners. Similarly, Akinola et al. [91] developed two QSARs based on TR binding data for sixty-eight mono-hydroxylated PCBs. While the ADs of these models were not formally defined, it is reasonable to assume that they were implicitly limited to these specific chemical classes due to their relatively small number and well-defined congeners. In the publication by de Lomana et al. [119], nine different QSARs were developed without defining a priori their ADs. Instead, ADs were assessed post hoc in terms of the Tanimoto coefficient by comparing the chemical space covered by the training sets with the chemical spaces covered by well-known datasets of pesticides, cosmetics, and drugs.

Despite its critical importance for ensuring the reliability of predictions for new chemicals, the AD was explicitly defined for fifty-four QSARs (see Table S1). As illustrated in Figure 5.7, a variety of methodologies were used for the AD definition of both classification and regression QSARs, showing that some studies integrated

multiple approaches while others relied on a single method. While some methods were used more frequently, others appeared in only a single instance.

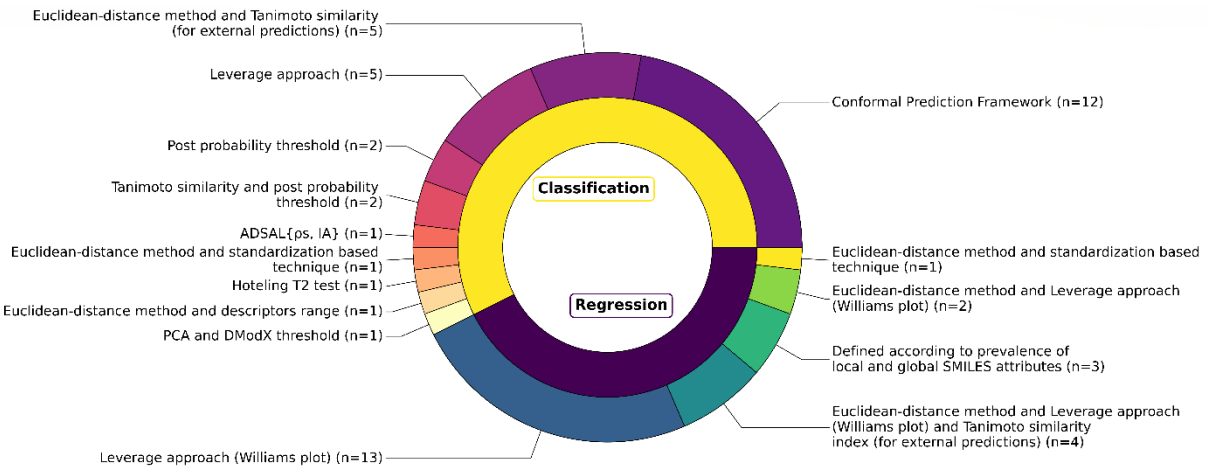

Figure 5.7. Types of AD definitions used in the selected QSAR models (and count).

The leverage approach was the most frequently used method, employed in approximately 44% of the QSAR models. This method was used either as a standalone technique [52,53,90,112,144,145,147,148] or in combination with other approaches [57,59,114], often complemented by the Williams plot as a graphical support for AD visualisation [52,53,57,59,90,112,114,144,147,148]. For example, in two different studies by Yang et al. [57,59] the leverage approach was combined with the Euclidean distance-based method to define the AD boundaries of two regression QSARs for TTR binding. In another study [114], Yang et al. used the same approach as before and included the Tanimoto similarity index to assess the reliability of external predictions for four regression QSARs for TTR binding; in the same study, they combined the Euclidean distance-based method with the Tanimoto similarity index to define the AD of five classification QSARs. The Euclidean distance method was additionally employed to define the AD of two QSARs developed by Kar et al. [133] and one QSAR by Kovarich et al. [146]. Kar et al. [133] combined it with the standardisation-based technique [173], while Kovarich et al. [146] combined it with the range of descriptor values in the training set. Zhang et al. [87] employed the Hotelling T2 test to measure the distance of new compounds from the centre of the training set in descriptor space in order to define the AD. Rybacka and colleagues [137] used a PCA to define the chemical space of the training set based on selected molecular descriptors, and then calculated the

distance-to-the-model (DModX) value for each compound. Methodologies less common than distance-based approaches were applied in six distinct publications [51,95,113,120,121,138]. Toropova et al. [138] defined the ADs of three QSARs according to the prevalence of local and global SMILES attributes in the training and validation sets, as proposed in their earlier publication [177]. Both Gadaleta et al. [120] and Rosenberg et al. [121] defined the ADs of their models in terms of the post probability of the predictions, with Rosenberg et al. [121] integrating the study of post probabilities with the Tanimoto similarity index. Liu et al. [95] characterised the AD in terms of weighted similarity density (ϱs) and weighted inconsistency of activities (IA) (AD_{SAL}{ϱs, IA}). Finally, both Dracheva et al. [51] and Sapounidou et al. [113] employed the CP framework to define the AD. As described in the studies, the CP quantifies the uncertainty of predictions by providing similarity scores, also termed as nonconformity scores, which can then be used to determine whether query compounds fall inside or outside the AD of a model.

Overall, a critical finding was the pronounced lack of QSAR models associated with a clearly defined AD. A clear definition of the AD is fundamental to increase confidence in the reliability of QSAR predictions and to accurately assess the degree of extrapolations. Without a defined AD, QSAR models risk being applied incorrectly and outside their intended scope, which can lead to the misuse of the tool and, ultimately, unreliable predictions.

5.3.8 Molecular descriptors: mechanistic interpretations and feature importance

Molecular descriptors encode for numerical representations of molecular structures and serve as independent variables in QSAR models. Thousands of molecular descriptors have been developed, reflecting the varied complexity of chemical structural representation. Molecular descriptors range from simple molecular properties (e.g., molecular weight (MW)) to highly complex ones (e.g., quantum chemical descriptors) [178]. Multiple types of software, either open or commercial, are available for their calculation [179].

Across the examined studies, an extensive range of molecular descriptors and software for their calculation was observed. The full list of software and molecular descriptors used for each model is provided in Table 5.2, where models are presented for each MIE to facilitate direct comparison. These descriptors encompassed multiple categories,

including physicochemical properties, FPs, constitutional, topological, electronic, and quantum chemical descriptors. It was a common practice to combine multiple software or libraries within a single study to compute different types of molecular descriptors.

Table 5.2. Summary of the molecular descriptors selected by each QSAR, grouped by MIE.

MIE	Ref.	Model ID	Chemical Class	Descriptors	Software
TTR	[53]	ID_5	Heterogeneous	AATSC1c; PubchemFP381; ATSC2s; nX	PaDEL [180]
		ID_6	Heterogeneous	naasC; SpMin4_Bhs; VE3_Dzs	PaDEL [180]
		ID_7	Heterogeneous	PubchemFP590; SpMax1_Bhe; PubchemFP18; GATS5c; AATSC1e; AATS4v	PaDEL [180]
	[51]	ID_16	Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[104]	ID_27	Heterogeneous	Calculation of extended fingerprints with a KNIME implementation of the CDK toolkit	CDK toolkit: https://cdk.github.io/
	[59]	ID_41	Halogenated phenols and thiophenols	logDOW(pH = 7.40); ω_{adj}; $dipole_{adj}$	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com); Gaussian 16; GsGrid 1.7 (http://gsgrid.codeplex.com)
	[114]	ID_42	Heterogeneous	V_{sadj}; Π_{adj}; μ_{adj}	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
		ID_43	Heterogeneous	V_{sadj}; O-059; μ_{adj}	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
		ID_44	Heterogeneous	V_{sadj}; H-050; nCbH	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
		ID_45	Heterogeneous	nArOH; V_{sadj}; ω_{adj}	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16;

				GsGrid 1.7, http://gsgrid.codeplex.com
	ID_46	Heterogeneous	V_{sadj}; C-024; nHDon	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
	ID_47	Heterogeneous	C-040; nCq; H-050; O-058; Πadj; O-056	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
	ID_48	Heterogeneous	log DOW(pH = 7.40); nArOH; O-057; nArNO2	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
	ID_49	Heterogeneous	$E_{HOMO-adj}$; nArOH; H052; ω_{adj}	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
	ID_50	Heterogeneous	log DOW(pH = 7.40); nArOH	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; GaussView 6.0; Gaussian 16; GsGrid 1.7, http://gsgrid.codeplex.com
[57]	ID_62	Phenolic DBPs	log D; $dipole_{adj}$	Marvin Sketch 15.6.29.0, 2015: ChemAxon, http://www.chemaxon.com; Gaussian 16
[133]	ID_66	PFCs	Me; nCsp2; H-050	DRAGON Version 6.0, 2011, http://www.talete.mi.it/
	ID_67	PFCs	IC3; ∑β's	DRAGON Version 6.0, 2011, http://www.talete.mi.it/
[87]	ID_70	Heterogeneous	Based on the following 14 molecular descriptors: TPSA; a_don; a_nOH; nX; PEOE_VSA_FNEG; PEOE_RPC-; density; PEOE_RPC+; diameter; PEOE_PC+; vsa_hyd; KierFlex; logP(o/w); opr_brigid	Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc.: Montreal, QC, Canada, 2015
[137]	ID_71	Heterogeneous	nArOH; nHDon; nCb-; nCRX3; nCH2RX; ALogPS_logP; nArOR; nCrq; nCq; nCp; nCs; nCbH	DRAGON version 6 [181].

	[144]	ID_78	PFCs and BFRs	nArOH; F03(Br..Br); HATS6m	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
		ID_79	PFCs and BFRs	R5u; F07[C-O]; nArOH	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
	[145]	ID_80	PFCs	AMW; HATS6m	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
		ID_81	PFCs	nH; HATS6m	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
		ID_82	PFCs	nH; F06[C-O]	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
		ID_83	PFCs	T(F..F); HATS6m	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2007
	[146]	ID_84	BFRs	DISPe; nArOH	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy, 2008
	[147]	ID_85	BFRs	qpmax; MATS6v	DRAGON Version 5.5 for Windows, Talete srl, Milan, Italy
TR α	[90]	ID_8	PFAS	X%; ICR	AlvaDesc [182]
	[104]	ID_28	Heterogeneous	Calculation of extended fingerprints with a KNIME implementation of the CDK toolkit	CDK toolkit: https://cdk.github.io/
TR β	[90]	ID_9	PFAS	X%; TPC	AlvaDesc [182]
	[51]	ID_17	Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
		ID_18	Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[104]	ID_29	Heterogeneous	Calculation of extended fingerprints with a KNIME implementation of the CDK toolkit	CDK toolkit: https://cdk.github.io/
	[112]	ID_39	PCNs	E_{LUMO}; ΔE; μ; Q_{xx}; Q_{yy}; Q_{yz}; q^+; $\log K_{ow}$; N_{Cl}; N_o	Gaussian 09 software.
	[113]	ID_40	Heterogeneous	Use of RDKit descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[122]	ID_63	PCBs	$\log K_{ow}$; ω; BER; nCl; EEig13d; JGI4	EPI Suite, version 4.1 (US EPA, 2012); DRAGON
		ID_64	PCBs	$\log K_{ow}$; ω; BER; nCl; EEig13d; JGI4	EPI Suite, version 4.1 (US EPA, 2012); DRAGON
	[138]	ID_72	Heterogeneous	Molecular optimal descriptor DCW(3, 10)	CORAL software: http://www.insilico.eu/coral

		ID_73	Heterogeneous	Molecular optimal descriptor DCW(1, 3)	CORAL software: http://www.insilico.eu/coral
		ID_74	Heterogeneous	Molecular optimal descriptor DCW(3, 4)	CORAL software: http://www.insilico.eu/coral
	[139]	ID_75	Heterogeneous	Thirty-five most statistically significant descriptors were identified: F04[N-Cl]; EEig03d; F06[C-Cl]; EEig08r; GATS7e; nArOH; EEig07r; EEig05d; EEig06d; TPSA(Tot); GGI1; BEHp4; SPI; C-026; ESpm01d; nCb-; Hy; GATS8v; T(O..O); BLTA96; IVDE; MATS1e; Ms; GATS6e; MATS6m; MATS5m; MATS2e; MATS1p; MATS8v; MATS6e; MATS8p; X4Av; X2Av; X0Av; Jhetp	Dragon software (version 5.4; Talete s.r.l., Milan, Italy)
		ID_76	Heterogeneous	Twenty-seven most statistically significant descriptors were identified: F08[C-Cl]; T(N..Cl); C-006; EEig06d; SEigm; ATS3m; ATS4m; BEHm6; T(O..Cl); ATS5m; ATS7m; BEHm7; Uindex; EEig04d; BELe3; EEig08d; HVcpx; PHI; BELm3; GGI8; BIC5; BEHml; JGI6; JGI7; BELml; GATS3p; VEA2	Dragon software (version 5.4; Talete s.r.l., Milan, Italy)
		ID_77	Heterogeneous	Thirty most statistically significant descriptors were identified: B05[O-O]; EEig03d; nArOH; GGI7; EEig05d; PW2; F04[C-N]; C-026; ESpm01d; AAC; GATS8p; Hy; PCR; GATS8v; F05[O-O]; O-057; MATS5v; IVDE; MATS1e; Ms; MATS5p; ARR; MATS5m; PHI; MATS8v; GATS1e; MATS8p; RBF; Jhetp; X1A	Dragon software (version 5.4; Talete s.r.l., Milan, Italy)
	[148]	ID_86	OH-PBDEs	nBr; logKow; I_A; E_{LUMO}; ω; μ^2	EPI Suite, version 4.0 (U.S. Environmental Protection Agency 2009); Gaussian 03 programs; DRAGON [181]
TR n.s.	[91]	ID_10	OH-PCBs	RDF35u; RDF55u; RDF85u; RDF65v	PaDEL [180]
		ID_11	OH-PCBs	RDF35u; RDF55u; RDF85u; RDF65v	PaDEL [180]
	[119]	ID_51	Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-	RDKit: Open-source cheminformatics. http://www.rdkit.org

				dimensional RDKit chemical descriptors	
	[123]	ID_65	PCBs; PBDEs	DELS; MAXDN; Mor31v; Ms; RDF040e; BER	DRAGON 5.5 for Windows, Talete srl, Milan, Italy, 2008
TSHR	[106]	ID_31	Heterogeneous	Thirty-nine descriptors were used, here sorted by their weight in descending order (top seven descriptors were used to build Model ID_32.): Sw < 0.1 mg/ml probability; LogSw; LogD(pH = 7.4); LogL; S; R2; E; LogS(pH = 7.4); logP; Solubility class; AAB/LogP; McGowan Volume; MW; Pi2; LogS(pH = 7.4)-; L; V; Sw < 1 mg/ml probability; No Of H Donors; Acid_pKa; LogSwLo; Sw > 10 mg/ml probability; Abraham's Alfa; NoOfRotBonds; A; Bo; 0Form; B; Form+; No Of H Acceptors; LogSwHi; Rel_pKa_ac; Base_pKa; Abraham's BetaH; Ertl TPSA; Form-; Rule of 5; Rel_pKa_bs; Form±	KOWWIN program (EPI Suite version 4.1.1, https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) to calculate logKow. Software for the calculation of the other molecular descriptors was not specified
		ID_32		Sw < 0.1 mg/ml probability; LogSw; LogD(pH = 7.4); LogL; S; R2; E	KOWWIN program (EPI Suite version 4.1.1, https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) to calculate logKow. Software for the calculation of the other molecular descriptors was not specified
		ID_33		The use of thirty-nine descriptors was reported in the study	KOWWIN program (EPI Suite version 4.1.1, https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) to calculate logKow. Software for the calculation of the other molecular descriptors was not specified
		ID_34		LogS, LogP, E	KOWWIN program (EPI Suite version 4.1.1, https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) to calculate logKow. Software for the calculation of the other molecular descriptors was not specified
		ID_35		Forty-one descriptors were used, here sorted by their weight in descending order:	KOWWIN program (EPI Suite version 4.1.1, https://www.epa.gov/tsca-

		Base_pKa; V; Abraham's Alfa; 0Form; AAB/LogP; CDocker Energy; NoOfRotBonds; S; LogSwLo; LogSwHi; CDocker Interaction Energy; Rel_pKa_bs; R2; E; LogD(pH = 7.4); LogS(pH = 7.4)-; Sw < 0.1 mg/ml probability; A; Sw > 10 mg/ml probability; Ertl TPSA; MW; logP; LogSw; Pi2; Abraham's BetaH; Solubility class; B; LogL; Sw < 1 mg/ml probability; L; Acid_pKa; Rel_pKa_ac; No Of H Acceptors; Bo; No Of H Donors; McGowan Volume; LogS(pH = 7.4); Form+; Form-; Form±; Rule of 5	screening-tools/epi-suitetm-estimation-program-interface) to calculate logKow. Software for the calculation of the other molecular descriptors was not specified
[51]	ID_19 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	ID_20 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
[119]	ID_52 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	ID_53 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
[95]	ID_15 Heterogeneous	Top twenty FPs with positive SHAP (Shapley additive explanation) values: PubchemFP12, PubchemFP259, PubchemFP257, PubchemFP256, PubchemFP628, PubchemFP185, PubchemFP258, PubchemFP2, PubchemFP143, PubchemFP146, PubchemFP656, PubchemFP633,	PaDEL [180]

			PubchemFP150, PubchemFP464, PubchemFP442, PubchemFP607, PubchemFP613, PubchemFP549, PubchemFP153, PubchemFP418	
TPO	[51]	ID_26 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[107]	ID_36 Heterogeneous	Use of Atom Pair Count (APC) fingerprints	PaDEL [180]
		ID_37 Heterogeneous	Use of Atom Pair Count (APC) fingerprints	PaDEL [180]
		ID_38 Heterogeneous	Use of Atom Pair Count (APC) fingerprints	PaDEL [180]
	[119]	ID_54 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[120]	ID_60 Heterogeneous	The top twenty ranked descriptors identified in the kNN model: GATS1e; NArOH; CATS2D_02_DL; MATS1e; MATS1s; C-026; CATS2D_03_DL; B10 [C-C]; MATS1m; 'SpMax2_Bh(s); MATS1p; nCb-; NX; Uc; 'P_VSA_i_1'; SpMAD_B(v); NCbH; GATS1s; MLOGP; Eta_C_A'	DRAGON v7.0.8., 2017: https://chm.kode-solutions.net/products_dragon.php
		ID_61 Heterogeneous	Based on 160 molecular descriptors	DRAGON v7.0.8., 2017: https://chm.kode-solutions.net/products_dragon.php
	[121]	ID_68 Heterogeneous	Based on scaffolds and structural features	Leadscope Predictive Data Miner (LPDM), Leadscope, Inc., (2016): http://www.leadscope.com/
		ID_69 Heterogeneous	The top ten most common structural features linked to active compounds: benzene, 1,3-dihydroxy-; Scaffold 288; benzene, 1-alkyl-,4-amino(NH2)-; benzene, 1,2-dihydroxy-; Scaffold 297; alcohol, alkenyl-; Scaffold 576; benzene, 1-alkoxy-,4-hydroxy-; Scaffold 306; Scaffold 574.	Leadscope Predictive Data Miner (LPDM), Leadscope, Inc., (2016): http://www.leadscope.com/

				The top ten most commons structural features linked to inactive compounds: Scaffold 110; Scaffold 342; Scaffold 210; Scaffold 253; Scaffold 303; Scaffold 108; benzene, 1-alkyl-,4-halo-; halide, benzyl-; Scaffold 454; Scaffold 194	
TBG	[52]	ID_1	PBBs	Molecular Weight (MW); Critical temperature (CT); Critical pressure (CP); Topological diameter (TD)	PaDEL [180]; Gaussian (Gaussian 09 (Gaussian Inc., Wallingford, CT, USA); ChemDraw 12.0
		ID_2	PBBs and OH-PBBs	Quadrupole moment Q_{yy} (Q_{yy}); Most negative Mulliken charge number (q-); Frequency (Freq); TD	PaDEL [180]; Gaussian (Gaussian 09 (Gaussian Inc., Wallingford, CT, USA); ChemDraw 12.0
		ID_3	PBBs and 2OH-PBBs	q^-; CP; TD; Topological Shape (TS)	PaDEL [180]; Gaussian (Gaussian 09 (Gaussian Inc., Wallingford, CT, USA); ChemDraw 12.0
		ID_4	PBBs, OH-PBBs, and 2OH-PBBs	q^-; CP; TD; CT	PaDEL [180]; Gaussian (Gaussian 09 (Gaussian Inc., Wallingford, CT, USA); ChemDraw 12.0
NIS	[51]	ID_25	Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[104]	ID_30	Heterogeneous	Calculation of extended fingerprints with a KNIME implementation of the CDK toolkit	CDK toolkit: https://cdk.github.io/
	[119]	ID_59	Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
Albumin	[93]	ID_12	PFAS	PDI; GATS8v; MATS8m; QED	AlvaDesc 2.0.16 [183]
		ID_13	PFAS	Eig12_AEA(bo); DECC; X4A	AlvaDesc 2.0.16 [183]
		ID_14	PFAS	QED; PDI; GATS8v; MATS8m	AlvaDesc 2.0.16 [183]
DIO1	[51]	ID_22	Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[119]	ID_56	Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org

DIO2	[51]	ID_23 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[119]	ID_57 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
DIO3	[51]	ID_24 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[119]	ID_58 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
TRHR	[51]	ID_21 Heterogeneous	Calculation of 119 RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org
	[119]	ID_55 Heterogeneous	Calculation of count-based Morgan fingerprints with a radius of 2 bonds and a length of 2048 bits, and of all 119 one-dimensional and two-dimensional RDKit chemical descriptors	RDKit: Open-source cheminformatics. http://www.rdkit.org

The mechanistic interpretation of a QSAR model is critically important because it allows for the identification of the chemical properties or structural features that most significantly contribute to the predicted endpoint, enhancing the scientific credibility and acceptance of predictions [157,158]. Furthermore, it can offer new insights into the molecular features driving the modelled activity, hence contributing to safe-by-design approach. However, mechanistic interpretation is not always straightforward. This is often due to the challenging interpretability of certain molecular descriptors or the complexity of the algorithms used in model development [184]. To overcome this, feature importance techniques are often employed to provide clarity and to pinpoint the most influential molecular descriptors among many, since not all descriptors contribute equally.

Mechanistic interpretation or the application of feature importance techniques was conducted for fifty-six QSARs selected in this review, accounting for approximately

65% of the total models. These approaches were applied across six specific MIEs: TTR, TR, TSHR, TPO, TBG, and albumin. The decision to conduct a straightforward mechanistic interpretation of selected molecular descriptors or to apply feature importance techniques was contingent upon various factors, including the type of modelling methodology employed, the chemical nature of the compounds modelled, and the specific types of molecular descriptors used. Relevant descriptors are influenced by the structural characteristics included in the dataset, which in turn depends on whether the dataset is composed of structurally heterogeneous chemicals or of compounds from a single chemical class.

Interpreted QSARs for TTR binding were either based on heterogenous organic chemicals or specific chemical classes, including halogenated phenols and thiophenols, PFAS and/or PFCs, and PBDEs and their hydroxylated metabolites. A strong consensus on the fundamental molecular properties influencing TTR binding was revealed, although different descriptors were used to represent those properties, highlighting the fact that various computational methods can effectively encode the same critical structural information. The most significant and consistently identified structural features were aromatic rings, halogen atoms, and hydroxyl groups. Examples of descriptors encoding for these features were nArOH (number of aromatic hydroxyls) and nX (number of halogen atoms), which consistently showed a positive correlation with TTR binding affinity. These can be referred to as "structural alerts", as their presence recalls the chemical structure of THs like T3 and T4. In addition, hydrophobicity was consistently recognised as a critical property driving TTR binding. Descriptors encoding for this property, such as logP and log DOW (pH = 7.40), were repeatedly selected in various QSARs. The hydrophobic nature of the TTR binding site for T4 justifies this observation [185]. Furthermore, descriptors like a_don, nHDon, and H-050 were selected to encode for hydrogen bond donor capacity, thereby emphasising the role of noncovalent interactions, such as hydrogen bonding and electrostatic interactions, between ligands and TTR. Furthermore, a consensus on the most significant features determining the TTR binding by PFAS (or PFCs) was highlighted across studies addressing this class of chemicals. These were mainly represented by the carbon chain length, MW and dimension, and terminal functional groups. An intermediate carbon chain length was found to be optimal for TTR binding. This information was encoded by descriptors like HATS6m and F06[C-O]. The most active PFAS were found to have an MW between 300 and 500 g/mol, as captured by the AMW descriptor. HATS6m, which encodes for molecular shape and dimension, was used to distinguish the activity of compounds with similar molecular weights.

F07[C-O] and nH were used to account for the presence of carboxylic or sulfonic acid terminal groups at a particular topological distance and to differentiate compounds based on their terminal functional group. As seen before, hydrophobicity was still recognised as a critical property driving TTR binding. A broad spectrum of molecular descriptors was used across these studies, encompassing quantum chemical and electronic descriptors, topological, structural, and constitutional ones, as well as functional group counts and logKOW. It was often observed that the same groups of descriptors were employed across different studies, especially when conducted by the same research groups. While all studies converged on similar key features for TTR binding, it is worth noting that the specific choice and subsequent interpretation of descriptors could be influenced by a research group's preferred modelling tools, their expertise, and their background. This implies that while the underlying findings may be consistent, their description might vary in the level of detail, depending on the specific approach adopted by the group.

As seen for TTR, interpreted QSARs for TR binding predictions were either based on heterogenous organic chemicals or specific classes. These included PCBs and their hydroxylated metabolites, PBDEs, PCNs, and PFAS. Similarly to TTR, TR modelling was performed using a wide array of molecular descriptors, and mechanistic interpretations were either more general or detailed. The presence and quantity of halogen atoms were consistently identified as being critical for TR binding. This information was encoded in descriptors like X% (percentage of halogen atoms) and nBr or nCl (number of bromine or chlorine atoms, respectively), which showed a positive correlation with TR binding affinity. In addition, molecular polarity was identified as a relevant property. Descriptors like EEig03d and EEig06d (edge adjacency indices weighted by dipole moments) and μ and μ2 (dipole moments) were positively correlated with TR binding, indicating that an increase in polarity could enhance affinity to TR. For PFAS, an optimal chain length was identified as a key determinant of TR binding, showing a moderate to high probability of binding for longer chains. Hydrophobicity was another key property consistently identified as a positive contributor to TR binding. Finally, electronic descriptors were selected to encode for the ability of a compound to accept or donate electrons and form hydrogen bonds with TR.

The interpretation of five QSARs for TPO inhibition by Seo et al. [107], Gadaleta et al. [120], and Rosenberg et al. [121] led to converged results, despite the use of different software to calculate molecular descriptors and different types of descriptors. The

presence of aromatic structures, either hydroxylated (e.g., phenols) or non-hydroxylated (e.g., anilines), and of various heteroatoms (including nitrogen, oxygen, sulphur, and halogens) were highlighted as key structural features for TPO inhibition. These findings pointed out how these structural features often mimic typical endogenous targets of TPO, like tyrosine residues, thereby exerting disrupting effects [120]. Additionally, the lipophilic nature of a compound was also identified as a critical property. Furthermore, valuable insights regarding typical structural features found in non-TPO inhibitors were provided [121], including ethers, esters, aryl halides, and tertiary amines. All of these findings offered a comprehensive picture of which structural features and/or properties either contribute to, or detract from, TPO inhibition.

Regarding TSHR, two QSARs by Liu et al. [95] and Xu et al. [106] were interpreted. According to Xu et al. [106], the inhibitory effect of compounds on the TSHR is primarily influenced by two key chemical descriptors: the probability of water solubility (encoded by the descriptor "Sw < 0.1 mg/mL probability") and lipophilicity (encoded by the descriptor "log D (pH = 7.4)"). The probability of water solubility was identified as the most influential factor because compounds must be able to diffuse through blood or body fluids to reach their biological target. Compounds with very low water solubility are less likely to be transported effectively, thus limiting their TSHR inhibitory potential. High lipophilicity was highlighted as key for TSHR inhibition, since this property describes the ability of compounds to penetrate the cell membrane and reach the transmembrane domain of TSHR. Nevertheless, several other molecular descriptors, reported in Table 5.2, were considered to account for factors influencing properties including dissociation properties, molecular flexibility, and electronic interactions. Liu et al. [95] employed the Shapley additive explanation (SHAP) technique to quantitatively assess the influence of each molecular feature, encoded as FPs, on TSHR agonism. While twenty different FPs were identified as having positive SHAP values, this analysis pointed out the contributions of lipophilicity, and aromatic and/or amino groups in promoting TSHR agonism.

TBG binding was only modelled by Yang et al., who developed four QSARs based on data for PBBs, including their mono-hydroxylated and di-hydroxylated metabolites [52]. The mechanistic interpretation indicated that hydroxylated metabolites exhibited a greater ability to bind with TBG, likely due to their capacity to establish hydrogen bonds or van der Waals interactions. Similarly, albumin binding was modelled only by Gallagher et al. [93], who developed three QSARs based on data for PFAS. Although

they only provided indications about the positive or negative contributions of the selected molecular descriptors based on their coefficients signs, the authors concluded that PFAS with chain lengths shorter than ten carbons demonstrated a higher albumin binding affinity compared with those with longer carbon chains.

Overall, although these findings provided valuable insights into the main structural features and properties that may cause TH system disruption, a drawback is the limited emphasis or, in some instances, the complete absence of mechanistic interpretations or the application of feature importance techniques. This may limit the confidence in QSAR models among both scientists and regulatory bodies. Therefore, considering the increasing demand for mechanistically informed NAMs to advance chemical hazard assessments, future research should prioritise and dedicate resources to improving the mechanistic understanding of QSAR models in order to promote wider acceptance and trust in these methodologies.

Finally, this review showed the wide array of software and molecular descriptors used in QSAR studies for TH system disruption, underscoring the dynamic nature of the field. The diversity in approaches to descriptor calculation and selection indicates a lack of a single, standardised tool or method. Instead, the choice of software and methodology appeared to be driven by factors such as the expertise of researchers, tool accessibility, and prior experience with specific platforms.

5.3.9 Recent advances

To keep this review up to date and to provide a picture of the field's evolution, this section was included to provide a concise picture of the key models published between January and July 2025. Table 5.3 and Table 5.4 are included for quick reference.

Charest et al. [187] developed a QSAR model for TTR binding prediction using RF on a dataset of 853 compounds. The AD was defined using prediction entropy (PS), a metric derived from the probability outputs from the RF. A core strength of their modelling process was the adoption of a "mechanistic a priori" approach. They first analysed crystal structures of TTR and performed docking studies of how chemicals bind to it, allowing the authors to select molecular descriptors that were known to be relevant to the binding mechanism. The chosen descriptors, obtained from the PaDEL descriptor library using the OPERA software v2.9 [192], included measures of hydrogen bonding (nHBacc, nHBDon), planarity (naAromAtom), and hydrophobicity (CrippenLogP), as well as more complex topological descriptors like ETA and ATSC to

capture fine structural details. They also included ZMIC descriptors to account for the structural specificity of protein binding. After training the model, they used the permutation importance and the mutual information methods to perform an a posteriori analysis to confirm the importance of their selected features. The authors found that the descriptors related to aromatic structures and hydrophobicity were highly relevant for TTR binding, which aligned with their initial mechanistic hypothesis and were consistent with mechanistic interpretations by previous studies described in Section 5.3.8. This combination of a priori and a posteriori analysis turned the model from a "black box" into a transparent and interpretable tool that is beyond the sole generation of predictions.

Table 5.3. Summary and main characteristics of relevant QSARs published from January to July 2025. C: classification-based; R: regression-based; Primary: data generated as part of the same study; Secondary: data collected from the existing literature.

Model ID	Reference	Year	MIE	Algorithm	C or R	Chemical Class	Data Source Type	Data Source Literature Reference(s)
ID_2025_1	[186]	2025	Albumin	MLR	R	Phenoxyacetic acid-derived congeners	Primary	[186]
ID_2025_2	[187]	2025	TTR	RF	C	Heterogenous	Secondary	[188]
ID_2025_3	[189]	2025	TTR	LDA	C	PFAS	Secondary	[190]
ID_2025_4	[189]	2025	TTR	MLR	R	PFAS	Secondary	[190]
ID_2025_5	[191]	2025	TTR	DTC	C	PFAS	Primary	[191]
ID_2025_6	[191]	2025	TTR	MLR	R	PFAS	Primary	[191]

Janicka et al. [186] developed a QSAR model designed to predict albumin binding based on data for twenty-nine phenoxyacetic acid-derived congeners. The authors first applied biopartitioning micellar chromatography (BMC) to derive an *in vitro* lipophilicity descriptor (logkBMC), which was used with other descriptors to develop an MLR-QSAR. The leverage approach, with the use of the Williams plot for graphical visualisation, was used to define the AD. Three key molecular descriptors defined the model: logkBMC to encode for lipophilicity, α to encode for polarizability, and the sum of hydrogen bond donors (HBD) and acceptors (HBA). Based on the descriptors' signs in the model's equation, binding to albumin was found to increase with higher lipophilicity and polarizability and to decrease with a greater number of hydrogen bond donors and acceptors. Two QSAR models were developed by Evangelista et al. [189] to predict the binding of PFAS to TTR, using a dataset of 134 PFAS. One

classification model was developed using LDA, while one regression model was developed using MLR. To ensure robustness and avoid overfitting, the models were subjected to a rigorous validation protocol including randomization procedures and leave-one-out bootstrapping. The AD was defined differently for each model. For the classification model, the AD was defined in terms of distance (cosine α) and post probabilities of classification. Shannon entropy was introduced to quantify the uncertainty associated with external predictions. For the regression model, the AD was defined in terms of the leverage approach, with the adoption of the Williams plot for graphical visualisation. A prediction interval was introduced to quantify the uncertainty associated with external predictions. The classification model was characterised by GATS3e, ATSC6p, GATS8m, and MIC2, while the regression model was characterised by piPC5, GGI9, and AATSC0e. The selected descriptors were consistent with prior *in vitro* and *in silico* (docking) findings regarding the major drivers of PFAS binding to TTR. The findings highlighted the importance of hydrogen bond formation and hydrophobic interactions to establish binding with TTR. The relevance of lipophilicity, molecular weight, and chain length were highlighted. The study performed by Sosnowska et al. [191] focused on the ability of PFAS to bind to TTR. The methodology involved the development of classification and regression QSAR models based on data from 45 PFAS. The classification model was developed using the decision tree classifier (DTC). Then, a single regression model was developed using MLR. The same algorithm was also used to perform a multiple regression model (MRM) approach, where a total of thirty-one single MLR-QSARs were developed from different data splits. The AD for the classification model was defined using a boundary box method, while the leverage approach with the support of the Williams plot was employed to define the AD of the regression QSARs. The classification model highlighted molecular size and structural complexity as key factors, using descriptors like SM4_D and GATS3m. The MLR model emphasised that compounds with heavier and more polar atoms tended to be more active, with descriptors like AMW and GATS7p. It also identified the importance of specific structural features, such as fluorine atoms at a topological distance of 10 bonds (B10[F-F]). Three molecular descriptors were frequently selected in the models developed through the MRM approach (JGI10, ATSC7c, and MATS6i), which further confirmed the importance of atomic charge and polarity. The QSAR models generated via the MRM approach were not included in this review for simplicity because their performance was comparable to that of the single model developed independently. The collective findings consistently showed that molecular size, complexity, and polarity were the primary

drivers of PFAS activity in disrupting TTR. Ultimately, it is highly relevant to cite the study by Cirino et al. [193]. Although their work did not propose any new QSAR models, its focus on optimising predictive performance through consensus modelling and evaluating the robustness of existing models made it a significant contribution to the field. Specifically, this study served as a retrospective analysis of computational strategies from the Tox24 Challenge [194], which aimed to advance computational toxicology for predicting chemical binding to TTR by using a large dataset of 1512 compounds [188]. The primary goal by Cirino et al. [193] was to analyse the models developed by the nine top-performing teams from the Tox24 Challenge and explore consensus strategies to enhance the predictive performances of single models. The participating teams adopted diverse strategies, with some relying on single-method models while others combined multiple approaches, such as descriptor-based and representation learning techniques. The study by Cirino et al. [193] demonstrated that consensus modelling improved the predictive accuracy for TTR binding, compared with individual models alone, achieving a lower error rate. Finally, an analysis to identify overrepresented functional groups in active compounds for TTR binding was performed. Unsurprisingly, groups similar to T4, such as phenols, aryl halides, and diarylethers, were highly frequent. Six other functional groups of potential concern were identified, including nitro compounds, arenes, and gem-trihalides.

Table 5.4. Summary of relevant QSARs published from January to July 2025 and selected molecular descriptors, grouped by MIE.

MIE	Ref. Model ID	Chemical class	Descriptors	Software
TTR	[187] ID_2025_2	Heterogenous	Thirty-one descriptors sorted by permutation importance: CrippenLogP; ATSC3c; ATSC5c; C1SP3; ETA_BetaP_s; naAromAtom; ZMIC1; ATSC4m; ZMIC5; ATSC4c; hmin; hmax; ATSC2m; ATSC5m; ETA_Beta_ns_d; ATSC0m; VE1_DzZ; C1SP2; ZMIC2; ATSC1m; nHBAcc; ZMIC3; ATSC3m; ATSC2c; ETA_dAlpha_A; ETA_Shape_Y; ATSC0c; maxdssC; ZMIC4; nHBDon; ATSC1c	PaDEL descriptors from OPERA software v2.9 [192]
	[189] ID_2025_3	PFAS	GATS3e; ATSC6p; GATS8m; MIC2	PaDEL [180]
	[189] ID_2025_4	PFAS	piPC5; GGI9; AATSC0e	PaDEL [180]
	[191] ID_2025_5	PFAS	SM4_D; GATS3m	AlvaDesc [182]
	[191] ID_2025_6	PFAS	AMW; GATS7p; B10[F-F]	AlvaDesc [182]

Albumin [186]	ID_2025_1	Phenoxyacetic acid-derived congeners	logkBMC; α; sum of HBD and HBA	ACD/Percepta software, version 1994–2012 (ACD/Labs, Advanced Chemistry Development, Inc., Toronto, ON, Canada)

5.4 Conclusions

This review highlighted the growing yet still-evolving landscape of QSAR models addressing MIEs leading to TH system disruption by chemical substances. While significant progress has been made, particularly due to the increased availability of HTS data, the field remains fragmented and challenges persist. This review highlighted a preference for classification-based models to predict categorical outcomes, instead of continuous toxicity values, and that, despite the rise in complex machine learning methods, simpler algorithms continued to be employed to leverage their interpretability and promote broader adoption.

This review revealed that modelling efforts were predominantly focused on key MIEs like TR and TTR, scarcely followed by TPO and TSHR. Critically, many other relevant MIEs, including the three deiodinases, NIS, TRHR, TBG, and albumin were significantly poorly addressed in QSAR research. A critical finding was the lack of QSAR modelling studies addressing MIEs related to DUOX, IYD, and pendrin inhibition, and those associated with cellular TH transport (specifically MCT8, MCT10, OATP1C1, OATP1A4, MDR1, and MRP2), highlighting critical areas for future investigations. Similarly, a limited number of chemical classes were addressed, leading to a very small number of local QSARs. Notably, while validation strategies were consistently employed, a critical finding was a frequent lack of explicitly defined ADs. Without clear AD definitions, QSARs risk being applied outside their scope, undermining decision-making confidence and leading to the incorrect use of QSARs. Furthermore, even though several types of molecular descriptors have been consistently identified as being relevant to model specific MIEs (e.g., TTR and TR), a limited emphasis on mechanistic interpretations was observed for many models, representing a critical drawback. However, the recent emergence of studies simultaneously covering multiple TH system-related endpoints demonstrated a growing awareness of the multifaceted nature of TH system disruption, offering a promising direction in aligning predictive modelling within AOP frameworks. The findings suggested a need for increased efforts in generating *in vitro* and *in silico* data for poorly addressed MIEs, broadening the chemical space of tested compounds, and ultimately developing new models. The successful application of integrated *in silico* approaches to generate activity data, such as molecular docking and dynamic simulations, has proven to be an effective strategy for developing QSARs when experimental data is limited or unavailable, presenting a valuable path forward for exploring multiple MIEs and for specific chemical classes. This would enable a more robust hazard assessment for entire groups of compounds. Future studies should

prioritise the development of QSAR models with clearly defined ADs and enhanced mechanistic interpretability to increase the reliability and transparency of and confidence in their predictions, ultimately to promote their wider acceptance as effective NAMs for TH system disruption assessment.

Acknowledgments and Funding: This research was funded by the PhD programme in Chemical and Environmental Sciences (DiSCA) at the University of Insubria; PhD scholarship awarded to Marco Evangelista.

Supporting information: Supplementary data associated with this article can be found in the online version at https://doi.org/10.3390/toxics13090799.

5.5 References

[1] Gore, A.C., La Merrill, M.A., Patisaul, H.B., Sargis, R., 2024. Endocrine Disrupting Chemicals: Threats to Human Health. The Endocrine Society and IPEN. ISBN # 978-1-955400-23-7, https://www.ipen.org/documents/endocrine-disrupting-chemicals-threats-human-health (last accessed on 06 October 2025).

[2] World Health Organisation (WHO), 2002. Damstra, T., Barlow, S., Bergman, A., Kavlock, R., Van Der Kraak, G., editors. Global assessment of the state-of-the-science of endocrine disruptors. Geneva: World Health Organization, https://www.who.int/publications/i/item/WHO-PSC-EDC-02.2 (last accessed on 06 October 2025).

[3] Ahn, C., Jeung, E.-B., 2023. Int. J. Mol. Sci. 24, 5342, https://doi.org/10.3390/ijms24065342.

[4] United Nations Environment Programme (UNEP), World Health Organisation (WHO), 2013. Bergman, A., Heindel, J.J., Jobling, S., Zoeller, R.T., editors. State of the science of endocrine disrupting chemicals – 2012. Geneva: United Nations Environment Programme, World Health Organization, https://www.who.int/publications/i/item/state-of-the-science-of-endocrine-disrupting-chemicals (last accessed on 06 October 2025).

[5] de Oliveira Santos, A.D., do Nascimento, M.T.L., da Silva de Freitas, A., Gomes de Carvalho, D., Bila, D.M., Hauser-Davis, R.A., Monteiro da Fonseca, E., Baptista Neto, J.A., 2023. Mar. Pollut. Bull. 197, 115727, https://doi.org/10.1016/j.marpolbul.2023.115727.

[6] Food and Agriculture Organisation of the United Nations (FAO), 2024. Exposure to endocrine disrupting chemicals – Changes from 2002 to 2024. Food Safety and Quality Series, No. 30, Rome, https://doi.org/10.4060/cd3005en.

[7] Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocr. Rev. 30, 293–342, https://doi.org/10.1210/er.2009-0002.

[8] Feldt-Rasmussen, U., Effraimidis, G., Klose, M., 2021. Mol. Cell. Endocrinol. 525, 111173, https://doi.org/10.1016/j.mce.2021.111173.

[9] Fekete, C., Lechan, R.M., 2014. Endocr. Rev. 35, 159–194, https://doi.org/10.1210/er.2013-1087.

[10] Sabatino, L., Vassalle, C., 2025. Biomolecules 15, 361, https://doi.org/10.3390/biom15030361.

[11] Mullur, R., Liu, Y.-Y., Brent, G.A., 2014. Physiol. Rev. 94, 355–382, https://doi.org/10.1152/physrev.00030.2013.

[12] De Luca, R., Davis, P.J., Lin, H.-Y., Gionfra, F., Percario, Z.A., Affabris, E., Pedersen, J.Z., Marchese, C., Trivedi, P., Anastasiadou, E., Negro, R., Incerpi, S., 2021. Front. Cell Dev. Biol. 8, 614030, https://doi.org/10.3389/fcell.2020.614030.

[13] Sawicka-Gutaj, N., Zawalna, N., Gut, P., Ruchała, M., 2022. Pharmacol. Rep. 74, 847–858, https://doi.org/10.1007/s43440-022-00377-w.

[14] Bassett, J.H.D., Williams, G.R., 2016. Endocr. Rev. 37, 135–187, https://doi.org/10.1210/er.2015-1106.

[15] Silva, J.F., Ocarino, N.M., Serakides, R., 2018. Biol. Reprod. 99, 907–921, https://doi.org/10.1093/biolre/ioy115.

[16] Yamakawa, H., Kato, T.S., Noh, J.Y., Yuasa, S., Kawamura, A., Fukuda, K., Aizawa, Y., 2021. Front. Physiol. 12, 606931, https://doi.org/10.3389/fphys.2021.606931.

[17] Brent, G.A., 2012. J. Clin. Invest. 122, 3035–3043, https://doi.org/10.1172/JCI60047.

[18] Alcaide Martin, A., Mayerl, S., 2023. Int. J. Mol. Sci. 24, 12352, https://doi.org/10.3390/ijms241512352.

[19] Giannocco, G., Kizys, M.M.L., Maciel, R.M., de Souza, J.S., 2021. Semin. Cell Dev. Biol. 114, 47–56, https://doi.org/10.1016/j.semcdb.2020.09.007.

[20] Adu-Gyamfi, E.A., Wang, Y.-X., Ding, Y.-B., 2020. Biol. Reprod. 102, 8–17, https://doi.org/10.1093/biolre/ioz182.

[21] Moog, N.K., Entringer, S., Heim, C., Wadhwa, P.D., Kathmann, N., Buss, C., 2017. Neuroscience 342, 68–100, https://doi.org/10.1016/j.neuroscience.2015.09.070.

[22] Köhrle, J., Frädrich, C., 2021. Best Pract. Res. Clin. Endocrinol. Metab. 35, 101562, https://doi.org/10.1016/j.beem.2021.101562.

[23] Kortenkamp, A., Axelstad, M., Baig, A.H., Bergman, Å., Bornehag, C.-G., Cenijn, P., Christiansen, S., Demeneix, B., Derakhshan, A., Fini, J.-B., Frädrich, C., Hamers, T., Hellwig, L., Köhrle, J., Korevaar, T.I.M., Lindberg, J., Martin, O., Meima, M.E., Mergenthaler, P., Nikolov, N., Du Pasquier, D., Peeters, R.P., Platzack, B., Ramhøj, L., Remaud, S., Renko, K., Scholze, M., Stachelscheid, H., Svingen, T., Wagenaars, F., Wedebye, E.B., Zoeller, R.T., 2020. Int. J. Mol. Sci. 21, 3123, https://doi.org/10.3390/ijms21093123.

[24] Oliveira, K.J., Chiamolera, M.I., Giannocco, G., Pazos-Moura, C.C., Ortiga-Carvalho, T.M., 2019. J. Mol. Endocrinol. 62, R1-R19, https://doi.org/10.1530/JME-18-0081.

[25] Calsolaro, V., Pasqualetti, G., Niccolai, F., Caraccio, N., Monzani, F., 2017. Int. J. Mol. Sci., 18, 2583 https://doi.org/10.3390/ijms18122583.

[26] Boas, M., Feldt-Rasmussen, U., Main, K.M., 2012. Mol. Cell. Endocrinol. 355, 240–248, https://doi.org/10.1016/j.mce.2011.09.005.

[27] Salazar, P., Villaseca, P., Cisternas, P., Inestrosa, N.C., 2021. Environ. Res. 200, 111345, https://doi.org/10.1016/j.envres.2021.111345.

[28] Alsen, M., Sinclair, C., Cooke, P., Ziadkhanpour, K., Genden, E., van Gerwen, M., 2021. Toxics 9, 1–26, https://doi.org/10.3390/toxics9010014.

[29] Olanrewaju, O.A., Asghar, R., Makwana, S., Yahya, M., Kumar, N., Khawar, M.H., Ahmed, A., Islam, T., Kumari, K., Shadmani, S., Ali, M., Kumar, S., Khatri, M., Varrassi, G., Mohamad, T., Olanrewaju, O.A., Asghar, R., Makwana, S., Yahya, M., Iv, N.B., Khawar, M.H., Ahmed, A., Islam, T., Kumari, K., Shadmani, S., Ali, M., Kumar, S., Khatri, M., Varrassi, G., Mohamad, T., 2024. Cureus J. Med. Sci. 16, e51574, https://doi.org/10.7759/cureus.51574.

[30] Brown, E.D.L., Obeng-Gyasi, B., Hall, J.E., Shekhar, S., 2023. Int. J. Mol. Sci. 24, 9815 https://doi.org/10.3390/ijms24129815.

[31] Murk, A.J., Rijntjes, E., Blaauboer, B.J., Clewell, R., Crofton, K.M., Dingemans, M.M.L., David Furlow, J., Kavlock, R., Köhrle, J., Opitz, R., Traas, T., Visser, T.J., Xia, M., Gutleb, A.C., 2013. Toxicol. in Vitro 27, 1320–1346, https://doi.org/10.1016/j.tiv.2013.02.012.

[32] Crofton, K.M., 2008. Int. J. Androl. 31, 209–223, https://doi.org/10.1111/j.1365-2605.2007.00857.x.

[33] Bernasconi, C., Sampani, S., Beronius, A., Coecke, S., Langezaal, I., Pistollato, F., Paini, A., Muñoz, A., Asturiol, D., Kienzler, A., Baron, G., Munn, S., Kandarova, H., Whelan, M., 2025. ALTEX, https://doi.org/10.14573/altex.2501152.

[34] European Commission, 2019. The European Green Deal. https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (last accessed on 06 October 2025).

[35] European Commission, 2020. Chemicals Strategy for Sustainability towards a Toxic-free Environment. https://environment.ec.europa.eu/strategy/chemicals-strategy_en (last accessed on 06 October 2025).

[36] Holmer, M.L., Holmberg, R.D., Despicht, C., Bouftas, N., Axelstad, M., Beronius, A., Zilliacus, J., Van Duursen, M., Svingen, T., 2025. Regul. Toxicol. Pharmacol. 162, 105883, https://doi.org/10.1016/j.yrtph.2025.105883.

[37] Ramhøj, L., Axelstad, M., Baert, Y., Cañas-Portilla, A.I., Chalmel, F., Dahmen, L., De La Vieja, A., Evrard, B., Haigis, A.-C., Hamers, T., Heikamp, K., Holbech, H., Iglesias-Hernandez, P., Knapen, D., Marchandise, L., Morthorst, J.E., Nikolov, N.G., Nissen, A.C.V.E., Oelgeschlaeger, M., Renko, K., Rogiers, V., Schüürmann, G., Stinckens, E., Stub, M.H., Torres-Ruiz, M., Van Duursen, M., Vanhaecke, T., Vergauwen, L., Wedebye, E.B., Svingen, T., 2023. Front. Toxicol. 5, 1189303 https://doi.org/10.3389/ftox.2023.1189303.

[38] European Chemicals Agency (ECHA), 2025. Key Areas of Regulatory Challenge. https://doi.org/10.2823/8572710.

[39] Organisation for Economic Co-operation and Development (OECD), 2018. Revised Guidance Document 150 on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. https://doi.org/10.1787/9789264304741-1-en.

[40] European Commission, 2017. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council (Text with EEA relevance.).

[41] European Commission, 2018. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting proper-ties (Text with EEA relevance.).

[42] European Commission, 2022. Commission Delegated Regulation (EU) 2023/707 of 19 December 2022 amending Regulation (EC) No 1272/2008 as regards hazard classes and criteria for the classification, labelling and pack-aging of substances and mixtures (Text with EEA relevance).

[43] Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., Villeneuve, D.L., 2010. Environ. Toxicol. Chem. 29, 730–741, https://doi.org/10.1002/etc.34.

[44] Svingen, T., Schwartz, C.L., Rosenmai, A.K., Ramhøj, L., Johansson, H.K.L., Hass, U., Draskau, M.K., Davidsen, N., Christiansen, S., Ballegaard, A.-S.R., Axelstad, M., 2022. Environ. Pollut. 304, 119242, https://doi.org/10.1016/j.envpol.2022.119242.

[45] Cronin, M.T.D., Richarz, A.-N., 2017. Appl. In Vitro Toxicol. 3, 286–297, https://doi.org/10.1089/aivt.2017.0021.

[46] Muratov, E.N., Bajorath, J., Sheridan, R.P., Tetko, I.V., Filimonov, D., Poroikov, V., Oprea, T.I., Baskin, I.I., Varnek, A., Roitberg, A., Isayev, O., Curtalolo, S., Fourches, D., Cohen, Y., Aspuru-Guzik, A., Winkler, D.A., Agrafiotis, D., Cherkasov, A., Tropsha, A., 2020. Chem. Soc. Rev. 49, 3525–3564, https://doi.org/10.1039/D0CS00098A.

[47] Noyes, P.D., Friedman, K.P., Browne, P., Haselman, J.T., Gilbert, M.E., Hornung, M.W., Barone, S., Crofton, K.M., Laws, S.C., Stoker, T.E., Simmons, S.O., Tietge, J.E., Degitz, S.J., 2019. Environ. Health Perspect. 127, 095001, https://doi.org/10.1289/EHP5297.

[48] Bernasconi, C., Langezaal, I., Bartnicka, J., Asturiol, D., Bowe, G., Coecke, S., Kienzler, A., Liska, R., Milcamps, A., Munoz-Pineiro, M.A., Pistollato, F., Whelan, M., 2023. Publications Office of the European Union, Luxembourg, https://doi.org/10.2760/862948.

[49] Sellami, A., Reau, M., Montes, M., Lagarde, N., 2022. Front. Endocrinol. 13, 986016, https://doi.org/10.3389/fendo.2022.986016.

[50] Vergauwen, L., Bajard, L., Tait, S., Langezaal, I., Sosnowska, A., Roncaglioni, A., Hessel, E., van den Brand, A.D., Haigis, A.-C., Novák, J., Hilscherová, K., Buławska, N., Papaioannou, N., Renieri, E., Spilioti, E., Spyropoulou, A., Gutleb, A.C., Holbech, H., Nikolopoulou, D., Jacobs, M.N., Knapen, D., 2024. Open. Res. Eur. 4, 242, https://doi.org/10.12688/openreseurope.18739.1.

[51] Dracheva, E., Norinder, U., Rydén, P., Engelhardt, J., Weiss, J.M., Andersson, P.L., 2022. Environ. Sci. Technol. 56, 8363–8372, https://doi.org/10.1021/acs.est.1c07762.

[52] Yang, L., Sun, P., Tao, L., Zhao, X., 2024. Chem. Biol. Interact. 397, 111075, https://doi.org/10.1016/j.cbi.2024.111075.

[53] Evangelista, M., Chirico, N., Papa, E., 2024. J. Hazard Mater. 480, 136188, https://doi.org/10.1016/j.jhazmat.2024.136188.

[54] Cao, J., Lin, Y., Guo, L.-H., Zhang, A.-Q., Wei, Y., Yang, Y., 2010. Toxicology 277, 20–28, https://doi.org/10.1016/j.tox.2010.08.012.

[55] Montaño, M., Cocco, E., Guignard, C., Marsh, G., Hoffmann, L., Bergman, Å., Gutleb, A.C., Murk, A.J., 2012. Toxicol. Sci. 130, 94–105, https://doi.org/10.1093/toxsci/kfs228.

[56] Grimm, F.A., Lehmler, H.-J., He, X., Robertson, L.W., Duffel, M.W., 2013. Environ. Health Perspect. 121, 657–662, https://doi.org/10.1289/ehp.1206198.

[57] Yang, X., Ou, W., Xi, Y., Chen, J., Liu, H., 2019. Environ. Sci. Technol. 53, 7019–7028, https://doi.org/10.1021/acs.est.9b00218.

[58] Xi, Y., Yang, X., Zhang, H., Liu, H., Watson, P., Yang, F., 2020. Chemosphere 242, 125135, https://doi.org/10.1016/j.chemosphere.2019.125135.

[59] Yang, X., Ou, W., Zhao, S., Wang, L., Chen, J., Kusko, R., Hong, H., Liu, H., 2021. Chemosphere 280, 130627, https://doi.org/10.1016/j.chemosphere.2021.130627.

[60] Rosenmai, A.K., Winge, S.B., Möller, M., Lundqvist, J., Wedebye, E.B., Nikolov, N.G., Lilith Johansson, H.K., Vinggaard, A.M., 2021. Chemosphere 263, 127703, https://doi.org/10.1016/j.chemosphere.2020.127703.

[61] Ren, X.M., Guo, L.-H., 2012. Environ. Sci. Technol. 46, 4633–4640, https://doi.org/10.1021/es2046074.

[62] Ren, X.-M., Qin, W.-P., Cao, L.-Y., Zhang, J., Yang, Y., Wan, B., Guo, L.-H., 2016. Toxicology 366–367, 32–42, https://doi.org/10.1016/j.tox.2016.08.011.

[63] Ouyang, X., Froment, J., Leonards, P.E.G., Christensen, G., Tollefsen, K.-E., de Boer, J., Thomas, K.V., Lamoree, M.H., 2017. Chemosphere 171, 722–728, https://doi.org/10.1016/j.chemosphere.2016.12.119.

[64] Qin, W.-P., Li, C.-H., Guo, L.-H., Ren, X.-M., Zhang, J.-Q., 2019. Environ. Sci. Process. Impacts 21, 950–956, https://doi.org/10.1039/C9EM00095J.

[65] Ren, X.-M., Yao, L., Xue, Q., Shi, J., Zhang, Q., Wang, P., Fu, J., Zhang, A., Qu, G., Jiang, G., 2020. Environ. Health Perspect. 128, 107008, https://doi.org/10.1289/EHP6498.

[66] Huang, K., Wang, X., Zhang, H., Zeng, L., Zhang, X., Wang, B., Zhou, Y., Jing, T., 2020. Environ. Sci. Technol. 54, 5437–5445, https://doi.org/10.1021/acs.est.9b05761.

[67] Hamers, T., Kortenkamp, A., Scholze, M., Molenaar, D., Cenijn, P.H., Weiss, J.M., 2020. Environ. Health Perspect. 128, 017015, https://doi.org/10.1289/EHP5911.

[68] van den Berg, K.J., 1990. Chem. Biol. Interact. 76, 63–75, https://doi.org/10.1016/0009-2797(90)90034-K.

[69] den Besten, C., Vet, J.J.R.M., Besselink, H.T., Kiel, G.S., van Berkel, B.J.M., Beems, R., van Bladeren, P.J., 1991. Toxicol. Appl. Pharmacol. 111, 69–81, https://doi.org/10.1016/0041-008X(91)90135-2.

[70] Lans, M.C., Klasson-Wehler, E., Willemsen, M., Meussen, E., Safe, S., Brouwer, A., 1993. Chem. Biol. Interact. 88, 7–21, https://doi.org/10.1016/0009-2797(93)90081-9.

[71] Cheek, A.O., Kow, K., Chen, J., McLachlan, J.A., 1999. Environ. Health Perspect. 107, 273–278, https://doi.org/10.1289/ehp.99107273.

[72] Meerts, I.A.T.M., van Zanden, J.J., Luijks, E.A.C., van Leeuwen-Bol, I., Marsh, G., Jakobsson, E., Bergman, Å., Brouwer, A., 2000. Toxicol. Sci. 56, 95–104, https://doi.org/10.1093/toxsci/56.1.95.

[73] Sandau, C.D., Meerts, I.A.T.M., Letcher, R.J., McAlees, A.J., Chittim, B., Brouwer, A., Norstrom, R.J., 2000. Environ. Sci. Technol. 34, 3871–3877, https://doi.org/10.1021/es001134f.

[74] Chauhan, K.R., Kodavanti, P.R.S., McKinney, J.D., 2000. Toxicol. Appl. Pharmacol. 162, 10–21, https://doi.org/10.1006/taap.1999.8826.

[75] Legler, J., Cenijn, P.H., Malmberg, T., Bergman, A., Brouwer, A., 2002. Organohalogen Compd. 53–56, https://hdl.handle.net/1871.1/77c55c42-b5ae-4874-a14e-363929dcb267.

[76] Meerts, I.A.T.M., Assink, Y., Cenijn, P.H., van den Berg, J.H.J., Weijers, B.M., Bergman, Å., Koeman, J.H., Brouwer, A., 2002. Toxicol. Sci. 68, 361–371, https://doi.org/10.1093/toxsci/68.2.361.

[77] Maia, F., Almeida, M. do R., Gales, L., Kijjoa, A., Pinto, M.M.M., Saraiva, M.J., Damas, A.M., 2005. Biochem. Pharmacol. 70, 1861–1869, https://doi.org/10.1016/j.bcp.2005.09.012.

[78] Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Kester, M.H.A., Andersson, P.L., Legler, J., Brouwer, A., 2006. Toxicol. Sci. 92, 157–173, https://doi.org/10.1093/toxsci/kfj187.

[79] Harju, M., Hamers, T., Kamstra, J.H., Sonneveld, E., Boon, J.P., Tysklind, M., Andersson, P.L., 2007. Environ. Toxicol. Chem. 26, 816–826, https://doi.org/10.1897/06-308R.1.

[80] Hamers, T., Kamstra, J.H., Sonneveld, E., Murk, A.J., Visser, T.J., Van Velzen, M.J.M., Brouwer, A., Bergman, Å., 2008. Mol. Nutr. Food Res. 52, 284–298, https://doi.org/10.1002/mnfr.200700104.

[81] Gales, L., Almeida, M.R., Arsequell, G., Valencia, G., Saraiva, M.J., Damas, A.M., 2008. Biochim. Biophis. Acta. 1784, 512–517, https://doi.org/10.1016/j.bbapap.2007.11.014.

[82] Weiss, J.M., Andersson, P.L., Lamoree, M.H., Leonards, P.E.G., van Leeuwen, S.P.J., Hamers, T., 2009. Toxicol. Sci. 109, 206–216, https://doi.org/10.1093/toxsci/kfp055.

[83] Hamers, T., Kamstra, J.H., Cenijn, P.H., Pencikova, K., Palkova, L., Simeckova, P., Vondracek, J., Andersson, P.L., Stenberg, M., Machala, M., 2011. Toxicol. Sci. 121, 88–100, https://doi.org/10.1093/toxsci/kfr043.

[84] Simon, E., Bytingsvik, J., Jonker, W., Leonards, P.E.G., de Boer, J., Jenssen, B.M., Lie, E., Aars, J., Hamers, T., Lamoree, M.H., 2011. Environ. Sci. Technol. 45, 7936–7944, https://doi.org/10.1021/es2016389.

[85] Simon, E., van Velzen, M., Brandsma, S.H., Lie, E., Løken, K., de Boer, J., Bytingsvik, J., Jenssen, B.M., Aars, J., Hamers, T., Lamoree, M.H., 2013. Environ. Sci. Technol. 47, 8902–8912, https://doi.org/10.1021/es401696u.

[86] Viluksela, M., Heikkinen, P., Ven, L.T.M. van der, Rendel, F., Roos, R., Esteban, J., Korkalainen, M., Lensu, S., Miettinen, H.M., Savolainen, K., Sankari, S., Lilienthal, H., Adamsson, A., Toppari,

J., Herlin, M., Finnilä, M., Tuukkanen, J., Leslie, H.A., Hamers, T., Hamscher, G., Al-Anati, L., Stenius, U., Dervola, K.-S., Bogen, I.-L., Fonnum, F., Andersson, P.L., Schrenk, D., Halldin, K., Håkansson, H., 2014. PLoS One 9, e104639, https://doi.org/10.1371/journal.pone.0104639.

[87] Zhang, J., Kamstra, J.H., Ghorbanzadeh, M., Weiss, J.M., Hamers, T., Andersson, P.L., 2015. Environ. Sci. Technol. 49, 10099–10107, https://doi.org/10.1021/acs.est.5b01742.

[88] Weiss, J.M., Andersson, P.L., Zhang, J., Simon, E., Leonards, P.E.G., Hamers, T., Lamoree, M.H., 2015. Anal. Bioanal. Chem. 407, 5625–5634, https://doi.org/10.1007/s00216-015-8736-9.

[89] Hill, K.L., Mortensen, Å.-K., Teclechiel, D., Willmore, W.G., Sylte, I., Jenssen, B.M., Letcher, R.J., 2018. Environ. Sci. Technol. 52, 1533–1541, https://doi.org/10.1021/acs.est.7b04617.

[90] Kowalska, D., Sosnowska, A., Bulawska, N., Stępnik, M., Besselink, H., Behnisch, P., Puzyn, T., 2023. Molecules 28, 479, https://doi.org/10.3390/molecules28020479.

[91] Akinola, L.K., Uzairu, A., Shallangwa, G.A., Abechi, S.E., 2023. SAR QSAR Environ. Res. 34, 267–284, https://doi.org/10.1080/1062936X.2023.2207039.

[92] Arulmozhiraja, S., Shiraishi, F., Okumura, T., Iida, M., Takigami, H., Edmonds, J.S., Morita, M., 2005. Toxicol. Sci. 84, 49–62, https://doi.org/10.1093/toxsci/kfi063.

[93] Gallagher, A., Kar, S., Sepúlveda, M.S., 2023. Molecules 28, 5375 https://doi.org/10.3390/molecules28145375.

[94] Jackson, T.W., Scheibly, C.M., Polera, M.E., Belcher, S.M., 2021. Environ. Sci. Technol. 55, 12291–12301, https://doi.org/10.1021/acs.est.1c01200.

[95] Liu, W., Wang, Z., Chen, J., Tang, W., Wang, H., 2023. Chem. Res. Toxicol. 36, 947–958, https://doi.org/10.1021/acs.chemrestox.3c00074.

[96] Neumann, S., Huang, W., Titus, S., Krause, G., Kleinau, G., Alberobello, A.T., Zheng, W., Southall, N.T., Inglese, J., Austin, C.P., Celi, F.S., Gavrilova, O., Thomas, C.J., Raaka, B.M., Gershengorn, M.C., 2009. Proc. Nat. Acad. Sci. 106, 12471–12476, https://doi.org/10.1073/pnas.0904506106.

[97] Jäschke, H., Neumann, S., Moore, S., Thomas, C.J., Colson, A.-O., Costanzi, S., Kleinau, G., Jiang, J.-K., Paschke, R., Raaka, B.M., Krause, G., Gershengorn, M.C., 2006. J. Biol. Chem. 281, 9841–9844, https://doi.org/10.1074/jbc.C600014200.

[98] Titus, S., Neumann, S., Zheng, W., Southall, N., Michael, S., Klumpp, C., Yasgar, A., Shinn, P., Thomas, C.J., Inglese, J., Gershengorn, M.C., Austin, C.P., 2008. SLAS Discov. 13, 120–127 https://doi.org/10.1177/1087057107313786.

[99] Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S.A., Attene-Ramos, M., Zhao, T., Austin, C.P., Simeonov, A., 2016. Nat. Commun. 7, 10425, https://doi.org/10.1038/ncomms10425.

[100] Huang, R., Xia, M., Sakamuru, S., Zhao, J., Lynch, C., Zhao, T., Zhu, H., Austin, C.P., Simeonov, A., 2018. Sci. Rep. 8, 3783, https://doi.org/10.1038/s41598-018-22046-w.

[101] Olker, J.H., Korte, J.J., Denny, J.S., Hartig, P.C., Cardon, M.C., Knutsen, C.N., Kent, P.M., Christensen, J.P., Degitz, S.J., Hornung, M.W., 2019. Toxicol. Sci. 168, 430–442, https://doi.org/10.1093/toxsci/kfy302.

[102] Wang, J., Hallinger, D.R., Murr, A.S., Buckalew, A.R., Lougee, R.R., Richard, A.M., Laws, S.C., Stoker, T.E., 2019. Environ. Int. 126, 377–386, https://doi.org/10.1016/j.envint.2019.02.024.

[103] Paul Friedman, K., Watt, E.D., Hornung, M.W., Hedge, J.M., Judson, R.S., Crofton, K.M., Houck, K.A., Simmons, S.O., 2016. Toxicol. Sci. 151, 160–180, https://doi.org/10.1093/toxsci/kfw034.

[104] Gadaleta, D., Spînu, N., Roncaglioni, A., Cronin, M.T.D., Benfenati, E., 2022. Int. J. Mol. Sci. 23, 3053 https://doi.org/10.3390/ijms23063053.

[105] Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L.J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M.P., Overington, J.P., Papadatos, G., Smit, I., Leach, A.R., 2017. Nucleic Acids Res. 45, D945–D954, https://doi.org/10.1093/nar/gkw1074.

[106] Xu, X., Wang, C., Gui, B., Yuan, X., Li, C., Zhao, Y., Martyniuk, C.J., Su, L., 2022. Environ. Res. 212, 113175, https://doi.org/10.1016/j.envres.2022.113175.

[107] Seo, M., Lim, C., Kwon, H., 2022. Food Sci. Biotechnol. 31, 483–495, https://doi.org/10.1007/s10068-022-01041-y.

[108] Carvalho, D.P., Ferreira, A.C.F., Coelho, S.M., Moraes, J.M., Camacho, M. a. S., Rosenthal, D., 2000. Braz. J. Med. Biol. Res. 33, 355–361, https://doi.org/10.1590/S0100-879X2000000300015.

[109] Divi, R.L., Doerge, D.R., 1996. Chem. Res. Toxicol. 9, 16–23, https://doi.org/10.1021/tx950076m.

[110] Habza-Kowalska, E., Kaczor, A.A., Żuk, J., Matosiuk, D., Gawlik-Dziki, U., 2019. Molecules 24, 2766 https://doi.org/10.3390/molecules24152766.

[111] Lee, J., 2015. Conversion of the organic breakdown products of glucosinolate to thiocyanate anions and their effects on thyroid hormone production. Ph.D. Thesis, Seoul National University, Seoul, Republic of Korea.

[112] Li, X., Gu, W., Zhang, B., Xin, X., Kang, Q., Yang, M., Chen, B., Li, Y., 2022. Environ. Int. 165, 107291, https://doi.org/10.1016/j.envint.2022.107291.

[113] Sapounidou, M., Norinder, U., Andersson, P.L., 2023. Chem. Res. Toxicol. 36, 53–65, https://doi.org/10.1021/acs.chemrestox.2c00267.

[114] Yang, X., Ou, W., Zhao, S., Xi, Y., Wang, L., Liu, H., 2021. ACS Sustain. Chem. Eng. 9, 5661–5672, https://doi.org/10.1021/acssuschemeng.1c00680.

[115] Van den Berg, K.J., van Raaij, J.A.G.M., Bragt, P.C., Notten, W.R.F., 1991. Arch. Toxicol. 65, 15–19, https://doi.org/10.1007/BF01973497.

[116] Marchesini, G.R., Meulenberg, E., Haasnoot, W., Mizuguchi, M., Irth, H., 2006. Anal. Chem. 78, 1107–1114, https://doi.org/10.1021/ac051399i.

[117] Marchesini, G.R., Meimaridou, A., Haasnoot, W., Meulenberg, E., Albertus, F., Mizuguchi, M., Takeuchi, M., Irth, H., Murk, A.J., 2008. Toxicol. Appl. Pharmacol. 232, 150–160, https://doi.org/10.1016/j.taap.2008.06.014.

[118] Purkey, H.E., Palaninathan, S.K., Kent, K.C., Smith, C., Safe, S.H., Sacchettini, J.C., Kelly, J.W., 2004. Chem. Biol. 11, 1719–1728, https://doi.org/10.1016/j.chembiol.2004.10.009.

[119] Garcia de Lomana, M., Weber, A.G., Birk, B., Landsiedel, R., Achenbach, J., Schleifer, K.-J., Mathea, M., Kirchmair, J., 2021. Chem. Res. Toxicol. 34, 396–411, https://doi.org/10.1021/acs.chemrestox.0c00304.

[120] Gadaleta, D., d'Alessandro, L., Marzo, M., Benfenati, E., Roncaglioni, A., 2021. Front. Pharmacol. 12 https://doi.org/10.3389/fphar.2021.713037.

[121] Rosenberg, S.A., Watt, E.D., Judson, R.S., Simmons, S.O., Paul Friedman, K., Dybdahl, M., Nikolov, N.G., Wedebye, E.B., 2017. Comput. Toxicol. 4, 11–21, https://doi.org/10.1016/j.comtox.2017.07.006.

[122] Bai, X., Yan, L., Ji, C., Zhang, Q., Dong, X., Chen, A., Zhao, M., 2018. Chemosphere 210, 312–319, https://doi.org/10.1016/j.chemosphere.2018.07.023.

[123] Yan, L., Zhang, Q., Huang, F., Nie, W.-W., Hu, C.-Q., Ying, H.-Z., Dong, X.-W., Zhao, M.-R., 2018. Chem. Phys. Lett. 706, 360–366, https://doi.org/10.1016/j.cplett.2018.06.022.

[124] Nakamura, N., Matsubara, K., Sanoh, S., Ohta, S., Uramaru, N., Kitamura, S., Yamaguchi, M., Sugihara, K., Fujimoto, N., 2013. Toxicol. Lett. 223, 192–197, https://doi.org/10.1016/j.toxlet.2013.09.007.

[125] Ren, X.-M., Guo, L.-H., 2013. Environ. Sci. Process. Impacts 15, 702–708, https://doi.org/10.1039/C3EM00023K.

[126] Amano, I., Miyazaki, W., Iwasaki, T., Shimokawa, N., Koibuchi, N., 2010. Ind. Health 48, 115–118, https://doi.org/10.2486/indhealth.48.115.

[127] Du, G., Shen, O., Sun, H., Fei, J., Lu, C., Song, L., Xia, Y., Wang, S., Wang, X., 2010. Toxicol. Sci. 116, 58–66, https://doi.org/10.1093/toxsci/kfq120.

[128] Li, J., Ma, M., Wang, Z., 2008. Environ. Toxicol. Chem. 27, 159–167, https://doi.org/10.1897/07-054.1.

[129] Sun, H., Shen, O.-X., Wang, X.-R., Zhou, L., Zhen, S., Chen, X., 2009. Toxicol. in Vitro 23, 950–954, https://doi.org/10.1016/j.tiv.2009.05.004.

[130] Hu, W., Liu, H., Sun, H., Shen, O., Wang, X., Lam, M.H.W., Giesy, J.P., Zhang, X., Yu, H., 2011. Mar. Pollut. Bull. 62, 2356–2361, https://doi.org/10.1016/j.marpolbul.2011.08.037.

[131] Liu, H., Hu, W., Sun, H., Shen, O., Wang, X., Lam, M.H.W., Giesy, J.P., Zhang, X., Yu, H., 2011. Mar. Pollut. Bull. 63, 287–296, https://doi.org/10.1016/j.marpolbul.2011.04.019.

[132] Kojima, H., Takeuchi, S., Uramaru, N., Sugihara, K., Yoshida, T., Kitamura, S., 2009. Environ. Health Perspect. 117, 1210–1218, https://doi.org/10.1289/ehp.0900753.

[133] Kar, S., Sepúlveda, M.S., Roy, K., Leszczynski, J., 2017. Chemosphere 184, 514–523, https://doi.org/10.1016/j.chemosphere.2017.06.024.

[134] Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., Kavlock, R.J., 2007. Toxicol. Sci. 95, 5–12, https://doi.org/10.1093/toxsci/kfl103.

[135] Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., Yang, C., Rathman, J., Martin, M.T., Wambaugh, J.F., Knudsen, T.B., Kancherla, J., Mansouri, K., Patlewicz, G., Williams, A.J., Little, S.B., Crofton, K.M., Thomas, R.S., 2016. Chem. Res. Toxicol. 29, 1225–1251, https://doi.org/10.1021/acs.chemrestox.6b00135.

[136] EDSP21 Work Plan, 2011. The Incorporation of In Silico Models and In Vitro High Throughput Assays in the Endocrine Disruptor Screening Program (EDSP) for Prioritization and Screening. https://www.epa.gov/sites/default/files/2015-07/documents/edsp21_work_plan_summary_overview_final.pdf (last accessed on 06 October 2025).

[137] Rybacka, A., Rudén, C., Tetko, I.V., Andersson, P.L., 2015. Chemosphere 139, 372–378, https://doi.org/10.1016/j.chemosphere.2015.07.036.

[138] Toropova, A.P., Toropov, A.A., Benfenati, E., 2015. Eur. J. Med. Chem. 101, 452–461, https://doi.org/10.1016/j.ejmech.2015.07.012.

[139] Politi, R., Rusyn, I., Tropsha, A., 2014. Toxicol. Appl. Pharmacol. 280, 177–189, https://doi.org/10.1016/j.taap.2014.07.009.

[140] Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J.P., 2012. Nucleic Acids Res. 40, D1100–D1107, https://doi.org/10.1093/nar/gkr777.

[141] Arnold, L.A., Kosinski, A., Estébanez-Perpiñá, E., Guy, R.K., 2007. J. Med. Chem. 50, 5269–5280, https://doi.org/10.1021/jm070556y.

[142] Hwang, J.Y., Arnold, L.A., Zhu, F., Kosinski, A., Mangano, T.J., Setola, V., Roth, B.L., Guy, R.K., 2009. J. Med. Chem. 52, 3892–3901, https://doi.org/10.1021/jm9002704.

[143] Hwang, J.Y., Attia, R.R., Zhu, F., Yang, L., Lemoff, A., Jeffries, C., Connelly, M.C., Guy, R.K., 2012. J. Med. Chem. 55, 2301–2310, https://doi.org/10.1021/jm201546m.

[144] Papa, E., Kovarich, S., Gramatica, P., 2013. SAR QSAR Environ. Res. 24, 333–349, https://doi.org/10.1080/1062936X.2013.773374.

[145] Kovarich, S., Papa, E., Li, J., Gramatica, P., 2012. SAR QSAR Environ. Res. 23, 207–220, https://doi.org/10.1080/1062936X.2012.657235.

[146] Kovarich, S., Papa, E., Gramatica, P., 2011. J. Hazard Mater. 190, 106–112, https://doi.org/10.1016/j.jhazmat.2011.03.008.

[147] Papa, E., Kovarich, S., Gramatica, P., 2010. Chem. Res. Toxicol. 23, 946–954, https://doi.org/10.1021/tx1000392.

[148] Li, F., Xie, Q., Li, X., Li, N., Chi, P., Chen, J., Wang, Z., Hao, C., 2010. Environ. Health Perspect. 118, 602–606, https://doi.org/10.1289/ehp.0901457.

[149] Organisation for Economic Co-operation and Development (OECD), 2014. New Scoping Document on in vitro and ex vivo Assays for the Identification of Modulators of Thyroid Hormone Signalling. https://doi.org/10.1787/9789264274716-en.

[150] Organisation for Economic Co-operation and Development (OECD), 2024. Thyroid in vitro methods: assessment reports by the thyroid disruption methods expert group: Reports assessing the validation status of assays from the EU-NETVAL activities. https://doi.org/10.1787/3786c75f-en.

[151] Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. Sci. Data 3, 160018, https://doi.org/10.1038/sdata.2016.18.

[152] Gilbert, M.E., O'Shaughnessy, K.L., Axelstad, M., 2020. Endocrinology 161, bqaa106, https://doi.org/10.1210/endocr/bqaa106.

[153] Pirhadi, S., Shiri, F., Ghasemi, J.B., 2015. RSC Adv. 5, 104635–104665, https://doi.org/10.1039/C5RA10729F.

[154] Schür, C., Gasser, L., Perez-Cruz, F., Schirmer, K., Baity-Jesi, M., 2023. Sci. Data 10, 718, https://doi.org/10.1038/s41597-023-02612-2.

[155] Schür, C., Schirmer, K., Baity-Jesi, M., 2025. Comput. Toxicol. 35, 100367, https://doi.org/10.1016/j.comtox.2025.100367.

[156] Wassenaar, P.N.H., Minnema, J., Vriend, J., Peijnenburg, W.J.G.M., Pennings, J.L.A., Kienhuis, A., 2024. Regul. Toxicol. Pharmacol. 148, 105589, https://doi.org/10.1016/j.yrtph.2024.105589.

[157] Organisation for Economic Co-operation and Development (OECD), 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en.

[158] Organisation for Economic Co-operation and Development (OECD), 2023. (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions. https://doi.org/10.1787/d96118f6-en.

[159] Tropsha, A., 2010. Mol. Inform. 29, 476–488, https://doi.org/10.1002/minf.201000061.

[160] Golbraikh, A., Tropsha, A., 2002. J. Mol. Graph. Model. 20, 269–276, https://doi.org/10.1016/S1093-3263(01)00123-1.

[161] Golbraikh, A., Tropsha, A., 2000. Mol. Divers. 5, 231–243, https://doi.org/10.1023/A:1021372108686.

[162] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second edition, Springer: New York, NY, USA, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[163] Todeschini, R., Consonni, V., Maiocchi, A., 1999. Chemometr. Intell. Lab. 46, 13–29, https://doi.org/10.1016/S0169-7439(98)00124-5.

[164] Tropsha, A., Gramatica, P., Gombar, V.K., 2003. QSAR Comb. Sci. 22, 69–77, https://doi.org/10.1002/qsar.200390007.

[165] Wehrens, R., Putter, H., Buydens, L.M.C., 2000. Chemometr. Intell. Lab. Syst. 54, 35–52, https://doi.org/10.1016/S0169-7439(00)00102-7.

[166] Ambure, P., Gajewicz-Skretna, A., Cordeiro, M.N.D.S., Roy, K., 2019. J. Chem. Inf. Model. 59, 4070–4076, https://doi.org/10.1021/acs.jcim.9b00476.

[167] Raste, S., Singh, R., Vaughan, J., Nair, V.N., 2022. arXiv, https://doi.org/10.48550/arXiv.2206.12353.

[168] Li, J., Zhao, T., Yang, Q., Du, S., Xu, L., 2025. Chemometr. Intell. Lab. Syst. 256, 105278, https://doi.org/10.1016/j.chemolab.2024.105278.

[169] Khan, A.A., 2022. Balanced Split: A new train-test data splitting strategy for imbalanced datasets. arXiv, https://doi.org/10.48550/arXiv.2212.11116.

[170] An, C., Park, Y.W., Ahn, S.S., Han, K., Kim, H., Lee, S.-K., 2021. PLOS ONE 16, e0256152, https://doi.org/10.1371/journal.pone.0256152.

[171] Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.-D., Lee, K.-H., Tropsha, A., 2003. J. Comput. Aided Mol. Des. 17, 241–253 https://doi.org/10.1023/A:1025386326946.

[172] Kennard, R.W., Stone, L.A., 1969. Technometrics 11, 137–148, https://doi.org/10.1080/00401706.1969.10490666.

[173] Roy, K., Kar, S., Ambure, P., 2015. Chemometr. Intell. Lab. Syst. 145, 22–29, https://doi.org/10.1016/j.chemolab.2015.04.013.

[174] Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klop-man, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.M., Tong, W., Veith, G., Yang, C., 2005. Altern. Lab. Anim. 33, 155–173, https://doi.org/10.1177/026119290503300209.

[175] Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., Todeschini, R., 2012. Molecules 17, 4791–4810, https://doi.org/10.3390/molecules17054791.

[176] Klingspohn, W., Mathea, M., ter Laak, A., Heinrich, N., Baumann, K., 2017. J. Cheminform. 9, 44, https://doi.org/10.1186/s13321-017-0230-2.

[177] Toropova, A.P., Toropov, A.A., Rallo, R., Leszczynska, D., Leszczynski, J., 2015. Ecotoxicol. Environ. Saf. 112, 39–45, https://doi.org/10.1016/j.ecoenv.2014.10.003.

[178] Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics. Wiley, https://doi.org/10.1002/9783527628766.ch20.

[179] Vasilev, B., Atanasova, M., 2025. Appl. Sci. 15, 1206, https://doi.org/10.3390/app15031206.

[180] Yap, C.W., 2011. J. Comput. Chem. 32, 1466–1474, https://doi.org/10.1002/jcc.21707.

[181] Todeschini R, Consonni V, 2000. Handbook of molecular descriptors. Wiley, https://doi.org/10.1002/9783527613106.

[182] Mauri, A., 2020. alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints. Ecotoxicological QSARs, 801–820 https://doi.org/10.1007/978-1-0716-0150-1_32.

[183] Mauri, A., Bertola, M., 2022. Int. J. Mol. Sci. 23, 12882. https://doi.org/10.3390/ijms232112882.

[184] Matveieva, M., Polishchuk, P., 2021. J. Cheminform. 13, 41, https://doi.org/10.1186/s13321-021-00519-x.

[185] Gião, T., Saavedra, J., Cotrina, E., Quintana, J., Llop, J., Arsequell, G., Cardoso, I., 2020. Int. J. Mol. Sci. 21, 2075 https://doi.org/10.3390/ijms21062075.

[186] Janicka, M., Sztanke, M., Sztanke, K., 2025. Molecules 30, 688, https://doi.org/10.3390/molecules30030688.

[187] Charest, N., Sinclair, G., Eytcheson, S.A., Chang, D.T., Martin, T.M., Lowe, C.N., Paul Friedman, K., Williams, A.J., 2025. J. Chem. Inf. Model. 65, 4426–4441, https://doi.org/10.1021/acs.jcim.5c00713.

[188] Eytcheson, S.A., Zosel, A.D., Olker, J.H., Hornung, M.W., Degitz, S.J., 2024. Chem. Res. Toxicol. 37, 1670–1681, https://doi.org/10.1021/acs.chemrestox.4c00215.

[189] Evangelista, M., Chirico, N., Papa, E., 2025. Toxics 13, 590, https://doi.org/10.3390/toxics13070590.

[190] Degitz, S.J., Olker, J.H., Denny, J.S., Degoey, P.P., Hartig, P.C., Cardon, M.C., Eytcheson, S.A., Haselman, J.T., Mayasich, S.A., Hornung, M.W., 2024. Toxicol. in Vitro 95, 105762, https://doi.org/10.1016/j.tiv.2023.105762.

[191] Sosnowska, A., Mudlaff, M., Mombelli, E., Behnisch, P., Zdybel, S., Besselink, H., Kuckelkorn, J., Bulawska, N., Kepka, K., Kowalska, D., Brouwer, A., Puzyn, T., 2025. J. Haz. Mater. 491, 137949, https://doi.org/10.1016/j.jhazmat.2025.137949.

[192] Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. J. Cheminform. 10, 10, https://doi.org/10.1186/s13321-018-0263-1.

[193] Cirino, T., Pinto, L., Iwan, M., Dougha, A., Lučić, B., Kraljević, A., Navoyan, Z., Tevosyan, A., Yeghiazaryan, H., Khondkaryan, L., Abelyan, N., Atoyan, V., Babayan, N., Iwashita, Y., Kimura, K., Komasaka, T., Shishido, K., Nakamura, T., Asada, M., Jain, S., Zakharov, A.V., Wang, H., Liu, W., Chupakhin, V., Uesawa, Y., 2025. Chem. Res. Toxicol. 38, 1061-1071, https://doi.org/10.1021/acs.chemrestox.5c00018.

[194] Tetko, I.V., 2024. Chem. Res. Toxicol. 37, 825–826, https://doi.org/10.1021/acs.chemrestox.4c00192.

CHAPTER 6

Contributions for Supporting the Regulatory Adoption of NAMs in the EU

The research and findings presented in this chapter draw heavily upon the data and methodologies presented as conference abstracts at SETAC Europe 35th Annual Meeting held in Vienna from 11 to 15 May, 2025, titled *Aquatic Toxicity Profile of Neurotoxic Substances* (1.07.P-Mo055) by Pia Talja, Marco Evangelista, Anna-Maija Nyman, Shoko Furuno, Ulla Simanainen, Kai Craenen and Marta Sobanska, and *In Silico Pre-Screening to Identify Neurotoxic Substances Which Are Considered Out of Applicability Domain for The Fish Embryo Acute Toxicity (FET) Test* (2.04.P-Tu186) by Ulla Simanainen, Marco Evangelista, Pia Talja, Anna-Maija Nyman, Shoko Furuno, Kai Craenen, Marta Sobanska, and Tomasz Sobanski. Abstract book available at: https://www.setac.org/resource/abstract-book-vienna-pdf.html (last accessed on 06 October 2025).

The content herein constitutes an original synthesis and critical re-elaboration of those conference publications, specifically adapted for the comprehensive scope of this dissertation.

6.1 Introduction

To evaluate the potential impact of a chemical substance on the aquatic ecosystem, acute and chronic toxicity are assessed using organisms from three distinct trophic levels: algae (primary producers), aquatic invertebrates (primary consumers), and fish (secondary consumers). This comprehensive approach is a cornerstone of European regulatory frameworks such as REACH [1], and it aims to determine a predicted no-effect concentration (PNEC) for environmental risk assessment (ERA) (see Section 1.1.2) and to classify substances as *hazardous to the aquatic environment* under the CLP regulation (Annex I, 4.1) [2]. In line with the overarching goal to move away from (vertebrate) animal testing for chemical hazard assessments and to phase out the most harmful chemicals, the use of the fish embryo toxicity (FET) test (OECD TG 236) [3] and of the fish cell line RTgill-W1 assay (OECD TG 249) [4] has been proposed as alternative to the traditional acute toxicity testing on fish (OECD TG 203) [5–8]. In this context, a proposed strategy involves using a combination of alternative-to-animal-testing methods such as QSAR models, the FET test, and the fish cell line assay, integrated under a weight of evidence (WoE) approach [7,8]. However, significant uncertainties exist regarding the regulatory application of this WoE approach for neurotoxic substances. Evidences suggest that the FET study may be less sensitive towards neurotoxicants than the traditional OECD TG 203 test, meaning that these substances may fall outside the applicability domain (AD) of the FET test [8]. This raises questions about how to ensure that alternative methods are sufficiently protective when evaluating these specific chemicals. Therefore, this work aimed to provide new insights to facilitate the implementation of new approach methodologies (NAMs) in place of the OECD TG 203 for assessing the acute environmental hazards of neurotoxic substances under REACH.

The first aim of this study was to analyse the aquatic toxicity profile of a reference list of neurotoxic substances to assess whether any aquatic trophic level would be most sensitive to neurotoxicants, and therefore would be used for classification purposes under the CLP regulation. As observed in a recent study [6], acute toxicity data on Daphnids (i.e., the typical model organism representing the trophic level of primary consumers) could reveal similar toxicity as fish for neurotoxic substances, and thus lead to similar or more stringent acute aquatic hazard classification for this chemical class, thereby impacting risk management measures.

The second aim was to evaluate whether existing QSAR models predicting potential neurotoxicity actually represent a feasible screening approach for the identification of neurotoxicants that might be out of the AD of the FET study.

6.2 Materials and methods

A list of 113 neurotoxic substances was used as reference of this work. This list was generated by Craenen and colleagues [9] for its use as a reference to aid in the development of NAMs for this type of chemical class. The authors included all the chemical substances having harmonised classification under the CLP regulation as specific target organ toxicity (STOT) after single and/or repeated exposure (STOT SE and STOT RE, respectively) 1 or 2 with the nervous system as a target organ, or STOT SE 3 based on narcotic effects.

For each chemical substance included in the reference list, data collection of acute aquatic toxicity data to fish, aquatic invertebrates, and algae was performed by consulting two main sources: harmonised classification and labelling (CLH) reports and the European Chemicals Agency (ECHA) chemicals database (ECHA CHEM, available at https://chem.echa.europa.eu/, last accessed on 06 October 2025). Available toxicity values were assessed for their reliability based on the information about the methodology employed for their generation (e.g., experimental test or predictive approach, type of test guideline, tested substance, etc.). Only reliable experimental toxicity values for the substance itself were included in the analysis. With the aim to identify the most sensitive trophic level and therefore which would drive the most stringent classification, the analysis was performed by comparing, for each chemical substance, the lowest acute toxicity value of each trophic level. A difference exceeding one order of magnitude between the lowest toxicity values for the two trophic levels of interest, i.e. Daphnids and fish, was used as a threshold to indicate a potential difference in the final classification of a substance.

To find tools capable of predicting both potential neurotoxicity on mammals and aquatic toxicity modes of action (MoA), a literature search was conducted. The search was focused on models, either statistical- or expert rule-based, that use molecular structures of chemicals to make predictions. For the purpose of this study, models that predict either neurotoxicity as apical endpoint or acetylcholinesterase (AChE) inhibition were considered. AChE inhibition was taken into account as it is a known mechanism leading to neurotoxicity and is often used as standalone evidence to conclude on neurotoxic potential of a compound. The following QSARs and/or predictive tools were selected and applied:

1. Assessment Tools for the Evaluation of Risk (ASTER) [10]: it is an application developed at the United States Environmental Protection Agency (US EPA) to

support chemical hazard assessment by providing either experimental or predicted data on physical-chemical properties and toxicities of chemicals. Among the others, ASTER includes several MoA related to potential neurotoxicity.

2. Linear discriminant analysis (LDA) models published by Martin and colleagues [11] to predict MoA related to potential neurotoxicity.
3. Derek Nexus v.6.3.0 [12]: this commercial tool allows for the detection of structural alerts associated to different levels of reasoning (e.g., probable, plausible) for the prediction of numerous toxicological endpoints, including potential neurotoxicity.
4. ProTox 3.0 [13]: it is a webserver including statistical-based models for the prediction of numerous toxicological endpoints, including potential neurotoxicity. Models provided binary predictions (i.e., active or inactive), each associated with a probability score for reliability assessment.

Their application was performed on 89 organic and mono-constituent neurotoxicants included in the reference list. Chemical abstract service registry numbers (CASRN) of neurotoxic substances were input in the US EPA CompTox Chemicals Dashboard [14] to download simplified molecular entry line system (SMILES) notations used to represent their chemical structures.

6.3 Results and discussion

Out of the 113 neurotoxic substances included in the reference list, 32 had reliable acute aquatic toxicity data available on all trophic levels, while 50 had reliable data on fish and Daphnids. Among the 32 neurotoxic substances analysed, no single trophic level was consistently identified as the most sensitive. Of these substances, fish and Daphnids were each the most sensitive trophic level for 34.4 % of the cases. For the remaining 31.2 %, algae proved to be the most sensitive. Data showed that sensitivity to these substances varied across the different trophic levels, with no data pattern emerging.

Then, an analysis of the most sensitive trophic level between fish and Daphnids was performed on the 50 neurotoxicants. Of these substances, fish was more sensitive than Daphnids for 26 neurotoxic substances, representing the 52%. However, for only one substance acute aquatic toxicity data on fish was more than one order of magnitude lower than acute aquatic toxicity data on *Daphnia magna*, while for another substance the difference was very close. This analysis suggested that for most of the cases the acute aquatic toxicity classification could be driven by the data on Daphnids. Although this result was limited to the reference list and did not take into account the entire universe of neurotoxic substances, it was in agreement with previous findings in the literature [6] where acute aquatic toxicity data on Daphnids proved to lead to the same classification as acute toxicity data on fish would for neurotoxicants.

Application of the predictive tools revealed their limited predictive capacity for potential neurotoxicity. Out of the 89 organic and mono-constituent neurotoxicants in the reference list, only 32, representing 36% of the total, were predicted as potential neurotoxicants by one or more models. However, only four neurotoxicants were predicted positive by more than one model, while the remaining 28 were predicted positive by only one model. All the remaining substances were predicted as negative. Furthermore, the capacity of the models to correctly identify neurotoxicants was influenced by the type of substance being analysed. Specific models, such as ASTER and LDA, were able to identify chemical substances registered for plant protection or biocidal products, suggesting that the predictive power of these tools may be highly dependent on a chemical's intended use or structural class.

6.4 Conclusions

Acute aquatic toxicity data on invertebrates (Daphnids) appeared sufficient to cover the acute aquatic hazards of neurotoxic substances. The analysis showed that even in cases where fish was more sensitive, the difference in toxicity values between fish and invertebrates was not significant enough to alter the conclusive hazard classification under CLP regulation. The findings of this study supported that the application of the FET test could serve as a viable alternative to traditional acute fish studies. For neurotoxic substances falling out the AD of the FET test, it was likely that the data on aquatic invertebrates would result in the same or even more stringent acute aquatic hazard classification. However, it is important to note that this conclusion was based on a limited dataset. More research is needed to confirm if data on aquatic invertebrates alone would consistently lead to the same classification for neurotoxic substances, especially for those outside the AD of the FET test.

The selected models had significant limitations as a screening method for identifying neurotoxicants that might fall outside the AD of the FET test. The tools demonstrated a low overall predictive capacity for potential neurotoxicity, with only a limited number of the investigated neurotoxicants being correctly identified by any single model. While combining predictions from multiple models provided a modest improvement, the models were still unable to identify a substantial portion of the neurotoxicants in the reference list. Furthermore, the predictive accuracy of the models was found to be dependent on the type of substance, with some tools providing positive predictions only for chemical substances registered for plant protection or biocidal uses. This indicated that these models may not be reliable for chemical classes employed in other uses. Given these limitations, future research should focus on developing more robust predictive models with a broader AD or specific for other types of substances, in order to appropriately integrate *in silico* approaches with other testing methods for a more reliable and comprehensive screening strategy for neurotoxicants.

6.5 References

[1] REACH Regulation (EC) No. 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals. Brussels, Belgium., 2006.

[2] European Commission, Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (Text with EEA relevance).

[3] Organisation for Economic Co-operation and Development (OECD), 2025. Test No. 236: Fish Embryo Acute Toxicity (FET) Test. https://doi.org/10.1787/9789264203709-en.

[4] Organisation for Economic Co-operation and Development (OECD), 2021. Test No. 249: Fish Cell Line Acute Toxicity - The RTgill-W1 cell line assay. https://doi.org/10.1787/c66d5190-en.

[5] Organisation for Economic Co-operation and Development (OECD), 2025. Test No. 203: Fish, Acute Toxicity Test. https://doi.org/10.1787/9789264069961-en.

[6] Schür, C., Paparella, M., Faßbender, C., Stoddart, G., Baity Jesi, M., Schirmer, K., 2025. Environ. Toxicol. Chem. 44, 2635–2647, https://doi.org/10.1093/etojnl/vgaf014.

[7] Mlnaříková, M., Pípal, M., Bláhová, L., Bláha, L., 2024. Environ. Sci. Eur. 36, 192, https://doi.org/10.1186/s12302-024-01015-3.

[8] Sobanska, M., Scholz, S., Nyman, A.-M., Cesnaitis, R., Gutierrez Alonso, S., Klüver, N., Kühne, R., Tyle, H., de Knecht, J., Dang, Z., Lundbergh, I., Carlon, C., De Coen, W., 2018. Environ. Toxicol. Chem. 37, 657–670, https://doi.org/10.1002/etc.4055.

[9] Craenen, K., Kosiaras, P., Hellsten, K., 2025. NeuroToxicology 109, 11–26, https://doi.org/10.1016/j.neuro.2025.05.006.

[10] Russom, C.L., Anderson, E.B., Greenwood, B.E., Pilli, A., 1991. Sci. Total Environ. 109–110, 667–670, https://doi.org/10.1016/0048-9697(91)90219-5.

[11] Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., Barron, M.G., 2013. J. Chem. Inf. Model. 53, 2229–2239, https://doi.org/10.1021/ci400267h.

[12] Derek Nexus, Lhasa Limited; https://www.lhasalimited.org/ (last accessed on 06 October 2025).

[13] Banerjee, P., Kemmler, E., Dunkel, M., Preissner, R., 2024. Nucleic Acids Res. 52, W513–W520, https://doi.org/10.1093/nar/gkae303.

[14] Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. J. Cheminform. 9, 61, https://doi.org/10.1186/s13321-017-0247-6.

CHAPTER 7

General Discussion, Conclusions, and Future Perspectives

The vast number of chemical substances in commerce poses a significant challenge to human and environmental health, contributing to what is known as the triple planetary crisis. Traditional toxicological and ecotoxicological assessments relying on (vertebrate) animal testing are no longer feasible for evaluating all chemical substances, given their high demand for resources and ethical considerations. This issue is particularly critical for endocrine disrupting chemicals (EDCs), a group of substances of global health concern and whose identification currently depends heavily on (vertebrate) animal-based studies.

This thesis directly addressed these challenges by developing and applying new quantitative structure-activity relationship (QSAR) models. The work aligned with current requirements on new approach methodologies (NAMs) to accelerate chemical hazard assessments and reduce reliance on (vertebrate) animal testing. The importance of NAMs is further emphasised at the international level, since the Organisation for Economic Co-operation and Development (OECD) has incorporated *in silico* approaches into its conceptual framework for the assessment of EDCs. In particular, the present project demonstrated the effectiveness of integrating QSAR models within the adverse outcome pathway (AOP) framework. This approach is essential for assessing EDCs and fulfilling current criteria for their identification, moving beyond simple hazard identification to a deeper, mechanistic understanding on how a chemical substance exerts its activity. Specifically, this research focused on endocrine disruption via interference with the thyroid hormone (TH) system. Novel QSAR models were developed to predict binding to human transthyretin (hTTR), a key molecular initiating event (MIE) in AOPs related to TH system disruption and neurotoxicity. The overarching aim of this research was to provide scientifically rigorous, transparent, and policy-relevant tools that can contribute in supporting the transition towards a toxic-free environment. Ultimately, the models developed in this work were applied to screen two classes of emerging contaminants, such as pharmaceutical and personal care products (PPCPs) and per- and polyfluoroalkyl substances (PFAS), to fill data gaps and prioritise them according to their TH system disrupting potential.

7.1 Development and application of new QSAR models

The core contribution of this thesis was the development of new QSAR models that fully adhere to the OECD guidelines and designed to fill data gaps and overcome challenges in identifying thyroid hormone system-disrupting chemicals (THSDCs).

In Chapter 2 (paper I in list of publications), a foundational work involved creating a new, comprehensive dataset by collecting and consolidating hTTR binding affinities and/or T4-hTTR competing potencies (expressed as relative competitive potency, or RP) from thirty-nine different literature sources. This dataset was documented as the largest and most comprehensive collection of its kind at that time and provided the bases for developing three new regression QSARs. These models were specifically calibrated using experimental data measured through three different *in vitro* assays: the 8-anilino-1-napthalenesulfonic acid (ANSA)-based binding assay, the fluorescence conjugate isothiocyanate (FITC)-T4-based binding assay, and the radiolabeled-T4 binding assay (RLBA). This multi-model approach directly addressed the inherent variability in different experimental methodologies. The models demonstrated strong statistical performance, with coefficient of determination (R^2) up to 0.89 and robust predictive ability on external test sets (Q^2_{F3} values up to 0.88, $RMSE_{TEST}$ values ranged from 0.39 to 0.66), affirming their quality and reliability. A key achievement of these models was their wider applicability domains (ADs) compared to similar previous QSARs, mainly due to the larger datasets utilised. This work also introduced the first-ever QSAR model based exclusively on data from the FITC-T4-based *in vitro* assay, as well as provided a model for the ANSA-based *in vitro* assay, both of which were timely contributions to the scientific and regulatory communities as these assays are currently under validation by the European Union Reference Laboratory for alternatives to animal testing (EURL ECVAM). The complete dataset underpinning this study, including both the quantitative data analysed and presented in Chapter 2, and the qualitative data collected but not included in this publication, has been deposited in the Zenodo repository. The dataset is publicly accessible and is listed as Dataset A in the datasets section of the present thesis. In particular, Dataset A comprises *in vitro* experimental hTTR binding affinity data of organic chemical substances collected from the literature. It is provided in Microsoft Excel format and includes a total of 734 individual data points corresponding to 497 distinct chemical substances. Each substance is reported primarily by its common name, since unique identifiers such as CASRN were not consistently available across the source literature and, in some cases,

were reported ambiguously or inconsistently. The data were systematically curated from 45 peer-reviewed scientific publications published up to and including 2021. For each entry, bibliographic information is provided in terms of main author, year of publication, and digital object identifier (DOI), ensuring traceability to the original sources. For each chemical substance, the dataset includes the *in vitro* experimental method used to measure hTTR binding. When available, quantitative measurements of T4–hTTR binding and chemical–hTTR binding are included together with their respective units of measurement. In addition, the RP value is reported, along with an explicit indication of whether this value was directly extracted from the literature or calculated within the dataset. Specifically, in cases where both T4–hTTR binding and chemical–hTTR binding values were available from the original studies, the RP was calculated and flagged accordingly. Conversely, when the literature reported only RP values without providing any of the underlying binding measurements, the RP was recorded as originally reported. In a limited number of cases, quantitative hTTR binding data were not available, and only qualitative information indicating the substance as active or inactive toward hTTR binding was provided. Overall, the aim of this curated dataset was to integrate the collected heterogeneous experimental evidence from the literature into a harmonized resource, supporting comparative analyses of chemical interactions with hTTR.

Chapter 4 (paper II in list of publications) expanded the modelling suite by developing models tailored for a specific high-priority class of chemicals, i.e., PFAS. This work resulted in a two-step methodology, where a classification QSAR was developed to first identify potential hTTR binders, and a regression QSAR to quantify their T4-hTTR competing potencies. This approach provided a powerful and pragmatic means to screen and prioritise large chemical libraries, moving from a qualitative assessment to a quantitative ranking. The models demonstrated high predictive power, with the classification model achieving training and test accuracies of 0.89 and 0.85, respectively, and the regression model showing an R^2 of 0.81 and Q^2_{F3} of 0.82. The QSAR presented in Chapter 4 demonstrated comparable or better performance than the few existing models for predicting hTTR disruption by PFAS (see Chapter 4 - Table 4.5 and Table 4.6), they featured significantly wider ADs, and were developed using non-commercial software and ANSA-based *in vitro* assay data. Rigorous validation procedures to check for overfitting and the presence of chance correlation confirmed the strong robustness of the proposed QSAR models. Their quality was further confirmed by validating the predictions obtained from the case study with experimental data from the literature. The complete dataset collected in this stage of

the study has been deposited in the Zenodo repository. The dataset is publicly accessible and is listed as Dataset B in the datasets section of the present thesis. Dataset B was constructed following the same data collection, curation, and annotation criteria adopted for Dataset A. In contrast to the broader chemical coverage of Dataset A, Dataset B comprises *in vitro* experimental hTTR binding affinity data of PFAS collected from the literature. It includes 391 individual entries corresponding to 276 distinct PFAS. The data were extracted and curated from 20 peer-reviewed scientific publications published up to June 2025. As for Dataset A, each entry is traceable to its original source through bibliographic information including main author, year of publication, and DOI.

Collectively, these models offered a systematic, rapid, and reproducible alternative to traditional animal-based testing, particularly for the early identification of potentially hazardous compounds. The two-step approach presented in Chapter 4 is highly valuable for prioritisation efforts and could be successfully replicated in other studies. Due to the limited size of the datasets used, and in light of the scarcity of QSAR models for predicting the potential TH system disruption (as illustrated in Chapter 5, paper III in list of publications), a focus on simpler, more transparent modelling methods was preferred over the use of more complex techniques. While the latter may promise higher predictive capacity, their "black box" nature often hinders understanding of how a prediction is obtained, undermining confidence and their adoption by scientists and regulatory authorities, who need to clearly trust a model and its outputs for critical applications. Hence, the use of these methods should be restricted to cases showing a much higher predictive capability, or when dealing with large and complex datasets. Furthermore, considering that QSAR research on TH system disruption is still in its early stage, as illustrated in Chapter 5, a simpler approach provides a more robust foundation for future developments and applications.

Finally, as illustrated in Chapter 6, the application of QSAR models for predicting (potential) neurotoxicity, which is one of the most critical adverse outcomes by TH system disruption, demonstrated that the current QSARs for this endpoint are still not adequate. This lack of predictive tools can slow and hinder the adoption of NAMs at the European regulatory level.

7.2 Enabling broader adoption of QSARs and screening of emerging contaminants

A central aspect of this thesis was the recognition that the utility of QSAR models is heavily dependent on their accessibility and ease of use. To ensure the models developed in this thesis could be widely adopted, they were implemented in the dedicated, open-source software, QSAR-ME Profiler. This decision to provide an open-access tool directly addressed a critical barrier to the broader adoption and application of predictive models, as many existing QSARs are tied to proprietary descriptors and/or software, limiting their widespread use. For many others, the parameters for their reproducibility are not always well documented. Through a user-friendly interface, the QSAR-ME Profiler not only makes the models freely available but also clearly quantifies both the AD and the uncertainties of predictions, thereby promoting transparency and reproducibility. This approach reduces the technical and financial barriers to apply the models, thus empowering a wider audience of scientists, regulators, and industry professionals to leverage the use of *in silico* toxicology and ecotoxicology for chemical hazard assessment. This aspect is strongly aligned with the major ambition of the Chemicals Strategy for Sustainability (CSS), where the early identification and regulation of (most hazardous) chemicals is imperative.

The practical utility of these QSARs was demonstrated in Chapter 3 (paper IV in list of publications) and Chapter 4, through two different case studies involving the screening of chemicals of emerging concerns, like PPCPs and PFAS. Screening of a subset of PPCPs by applying the QSARs presented in Chapter 2 allowed for the identification of concerning chemicals for their TH system disrupting potential. In addition, the combination of removal efficiency data with *in silico* predictions for multiple properties allowed for the prioritisation of 16 different PPCPs according to their potential to be PBT, PMT, and EDCs, warranting further investigation. The models presented in Chapter 4 were applied to the OECD List of PFAS, a comprehensive inventory of almost 5000 chemicals for which hazard information is largely unknown. This application successfully filled huge data gaps by providing a critical initial assessment of the potential to disrupt the TH system by thousands of PFAS. The screening efforts led to the identification and prioritisation of several structural categories of high concern, including “per-and polyfluoroalkyl ether-based compounds”, “perfluoroalkyl carbonyl compounds”, and “perfluoroalkane sulfonyl compounds”. A significant finding was that forty-nine out of the screened PFAS were predicted to have a stronger binding affinity for hTTR than the natural ligand T4. These

results were crucial because they can be used to move toward more efficient group-based strategies, as well as provided a rationale for PFAS prioritisation in future regulatory actions and risk management efforts. In addition, the reliability of the predictions was confirmed against existing experimental data found via a comprehensive literature search. Despite the scarcity of available data, the findings showed a high degree of agreement with the QSAR predictions.

7.3 Integrating QSAR modelling within a mechanistic framework

A central element of this thesis, emphasised across Chapter 2, Chapter 4, and Chapter 5, was the importance of moving beyond mere predictions to a deeper mechanistic understanding of chemical toxicity. The QSARs developed in this thesis were not just stand-alone computational tools, but were explicitly framed within the broader scientific structure of the AOP framework. The alignment of QSAR modelling with the AOP framework is crucial as it provides a scientifically robust foundation by clearly defining the molecular activity of a chemical and its potential for causing an adverse effect. This is particularly important for endocrine disruption assessment, as this integrated approach can generate the data necessary to recognise a chemical substance as an EDC according to the new criteria adopted under main European chemical regulations and the key characteristics outlined by experts in the field. This approach increases the likelihood that these models and their findings will be adopted by regulators and applied researchers.

In this thesis, the models were developed using simple and transparent modelling algorithms, such as multiple linear regression (MLR) and linear discriminant analysis (LDA), from which a direct link between selected molecular descriptors and the predicted biological mechanism may emerge. This might contrast to some more complex algorithms (e.g., artificial neural networks). However, as previously described, while preferred for handling vast amounts of data, they are often criticised for their "black box" nature and lack of transparency. Historically, this represents a major barrier to their adoption by regulatory bodies and the wider scientific community.

The use of simpler algorithms for QSARs development in Chapter 2 and 4, while still achieving good predictive performance, allowed for a more straightforward mechanistic interpretation of the selected molecular descriptors. This provided

valuable insights into the structural features driving hTTR binding. Across the different models presented in Chapter 2, a strong consensus emerged that binding affinity is correlated with structural features that mimic THs structures, particularly the presence of hydroxylated and aromatic rings. Regarding PFAS, the mechanistic interpretation of the QSAR models presented in Chapter 4 highlighted that hTTR binding is directly influenced by properties such as chain length, lipophilicity, and molecular weight, and is governed by hydrophobic interactions and hydrogen bond formation. These insights are essential for a variety of applications, such as informing green chemistry and safe-by-design initiatives aimed at developing safer chemical alternatives.

The inclusion of a comprehensive table in Chapter 5 (Table 5.2) listing the specific molecular descriptors used for each QSAR model for TH system disruption is a critical requirement for advancing this field. Consolidating this knowledge is mandatory to facilitate direct and meaningful comparisons across different independent studies, and ultimately to provide insights into the molecular basis of toxicity.

Although paramount for enhancing confidence and acceptance of model predictions among the scientific community and regulatory authorities, the mechanistic interpretation of QSARs is not always straightforward, often due to the challenging nature of interpreting complex molecular descriptors or modelling algorithms, as highlighted in Chapter 5. Due to these challenges, the strategic application of feature importance techniques and the emphasis on simple and interpretable models remain a crucial aspect of QSAR development. However, the emergence of powerful artificial intelligence and the spread of machine and deep learning methods cannot be ignored. Given the growing need for reliable alternatives to traditional animal testing and their proper implementation in the current regulatory frameworks, it is imperative to focus on advancing the interpretability of all of the models. This will allow to leverage their predictive power while simultaneously building trust necessary for their widespread regulatory adoption.

7.4 Data gaps and overcoming challenges

A foundational step in this thesis was the comprehensive data collection and curation, an effort that spanned Chapter 2, 4, and 5, and involved aggregating a vast amount of existing literature and data. This process of consolidating existing knowledge was not just a preliminary step for QSARs development, but a key finding of this project.

Indeed, it allowed for the systematic identification of knowledge gaps in the current state-of-the-art of QSAR modelling for TH system disruption. Results from the review work presented in Chapter 5 highlighted that past modelling efforts have been concentrated on a limited number of relevant MIEs for TH system disruption, particularly THs receptors (TRs) and TTR. This leaves other critical MIEs significantly underexplored or overlooked by QSAR researchers, including the three deiodinases (i.e., DIO1, DIO2, and DIO3), the sodium-iodide symporter (NIS), the thyrotropin releasing hormone receptor (TRHR), other THs distributor proteins (i.e., albumin and thyroid binding globulin), and those involved in THs cellular transport.

The analysis of the consolidated dataset presented in Chapter 2 not only allowed for the collection of available data, but further revealed that a number of chemical categories were critically underrepresented in the scientific literature. Although compounds such as some pesticides, bisphenol derivatives, fatty acids, phthalates, and phosphates have been experimentally demonstrated as THSDCs, many other compounds belonging to these same categories and other categories have not been studied. This is a critical observation, given that compounds with similar structure often exhibit similar activity and toxic potential. The analysis of the dataset composition enabled the identification of research gaps and lay the groundwork for future studies to focus on chemical classes of concern that were previously overlooked. For a more complete narrative, see dataset A.

A similar finding was highlighted in Chapter 5, where only a narrow range of chemical classes were addressed by existing models, such as such as polychlorinated biphenyls (PCBs), polybrominated diphenyl ethers (PBDEs), and polybrominated biphenyls (PBBs), leading to the availability of a small number of local QSARs. It is important to note that these chemicals are characterised by halogenated and aromatic structures, which have a structure resembling that of THs. Consequently, the current understanding of TH system-disrupting potential of chemicals with different structures is still very limited. This observation underscores the urgent need for a continuous and targeted effort to generate new data, in order to refine and expand the ADs of existing models and to enhance the development of both new global and local QSARs. This challenge was sought to be addressed by the dedicated PFAS models presented in Chapter 4. Nevertheless, despite the use of the largest homogeneous dataset of hTTR binding by PFAS for QSARs development, the analysis revealed that certain structural categories of PFAS, such as "other PFAA precursors or related compounds — semifluorinated", were still poorly covered by models ADs. This

resulted in many predictions for these categories falling outside the ADs of the models, highlighting a critical area for future data generation efforts for a high-priority class of chemicals.

To address these gaps, a concerted effort to generate new, high-quality data is essential. The development and application of new, and preferably validated, *in vitro* assays is a crucial step. While large-scale programs like Tox21 and ToxCast are already generating extensive high-throughput screening (HTS) data for a variety of endpoints, progress at an international level is slower. This is primarily because, at present, there are no fully validated *in vitro* assays for TH system-related endpoints, although the EURL ECVAM is actively working on that. The availability of such validated testing methods is crucial for ensuring that the data generated is of high quality and can be reliably used also for developing new QSAR models. This is because the fundamental principle of *garbage in, garbage out* applies directly to predictive modelling, underscoring that the reliability and predictive power of any model are directly dependent on the quality of its input data. Recognising these challenges, a strategy to mitigate the negative impact of using data from non-validated *in vitro* assays was adopted in this thesis. In Chapter 2, separate QSARs were developed using only data from the same *in vitro* assay to avoid the bias from mixing data from different assays. Secondly, the use of the RP was preferred over binding affinity as modelling endpoint in order to further reduce potential bias between data from different sources. Indeed, RP normalises binding affinity data against a common reference compound (i.e., T4), minimising variability stemming from different experimental conditions and laboratories. Building on this robust approach, it is recommended that future studies in predictive toxicology adopt similar strategies to ensure the highest possible data quality when they come from non-validated methods.

The lack of validated *in vitro* methodologies for TH system disruption assessment and the heterogenous nature of the data from the literature also require greater attention in evaluating the uncertainty associated with each model and each individual prediction to foster trust in and the wider adoption of QSARs. A central methodological advancement of this research, as described in Chapter 2 and Chapter 4, was the introduction of uncertainty quantification beyond the traditional definition of ADs. There is not a single and universally accepted methodology for AD definition, instead various approaches are employed, each with its own level of restrictiveness. On one hand, a less restrictive AD allows a model to be more broadly applicable, thereby including a wider range of chemical substances, but with potentially less

reliable predictions. On the other hand, a narrow and more restrictive AD offers greater confidence in the predictions that a model generates, but for a more limited number of compounds. In this thesis, metrics such as Shannon entropy for classification models, and prediction intervals for regression models, were introduced in addition to the AD definition, providing additional information to address reliability of predictions, aligning with the broader scientific and regulatory need for greater trust and confidence in predictive models. This assumes further importance given the frequent lack of defined ADs associated to QSAR models for TH system disruption, as described in Chapter 5.

Beyond the sole use of data generated through *in vitro* assays, multiple studies, as shown in Chapter 5, have convincingly demonstrated that integrating diverse *in silico* approaches is a highly effective strategy for generating valuable activity data, especially when experimental data is scarce or unavailable. Combining methods like molecular docking and molecular dynamics simulations allows for the quantification of molecular binding between chemical substances and biological targets at the molecular level. This powerful approach allows for the prediction of the potential activity and toxicity of compounds and the reduction of extensive and costly laboratory work. The data generated through these methods can then be used to develop new QSAR models, offering a practical and effective solution to bridge data gaps and accelerate chemical hazard assessment. Based on that, a critical recommendation for future research is to foster greater collaboration between experts in different *in silico* methods. Since each method has its own set of strengths and weaknesses, the full potential of these can be realised through a more integrated and interdisciplinary approach. The identification of the explicit gaps identified in this thesis is a crucial contribution in the field of TH system disruption assessment. It provides a clear, data-driven rational for directing future experimental and computational research towards these specific, underexplored areas. By highlighting where new data is most urgently needed, this work contributes to a more strategic and resource-efficient approach to advance the field of THSDCs assessment.

7.5 Future perspectives

The work presented in this thesis demonstrated that QSAR models are powerful alternatives for advancing chemical hazard assessment, even for the complex issue of TH system disruption. By developing robust and transparent models (Chapter 2 and

Chapter 4), this research has provided valuable tools for identifying, prioritising, and understanding the potential TH system disruption of a wide range of chemicals, including a specific focus on emerging contaminants like PFAS and PPCPs (Chapter 3 and Chapter 4). This research also provided a clear roadmap for the future by consolidating existing knowledge on QSAR modelling of MIEs relevant for TH system disruption (Chapter 5). The systematic identification of modelling gaps for specific MIEs and chemical classes is arguably as important as the models themselves, directing future efforts toward the most critical areas. Furthermore, the introduction of uncertainty quantification represented an additional and crucial step beyond the sole AD definition to build greater trust and confidence in QSARs. Finally, implementing the QSARs in the open-access QSAR-ME Profiler software was a practical contribution that lowers the barrier to adoption, ensuring that the insights gained from this thesis can be widely applied.

Looking ahead, to increase the potential of QSAR approaches for TH system disruption assessment and to promote their regulatory adoption, a collaborative and concerted effort by experts is needed to address the following key points:

- to generate targeted data by prioritising research on underexplored MIEs and chemical classes. The use of integrated *in silico* approaches (e.g., QSARs, molecular docking, molecular dynamic simulations) can be an effective strategy, especially for developing local QSARs, when experimental data is lacking.
- to advance methodological transparency by developing and refining methods for quantifying predictions uncertainty and by performing robust mechanistic interpretations. This is particularly important given the spread of complex machine learning approaches and the need to build trust for the regulatory adoption of NAMs.
- to promote the use of open-source software and adherence to FAIR (findable, accessible, interoperable, reusable) data principles to facilitate collaboration, reproducibility, and the widespread adoption of QSARs.
- to enhance the use of consensus modelling approaches to mitigate the inherent limitations of individual models. This strategy has proven effective in the field of endocrine disruption, as described in Chapter 5 and demonstrated by the COMPARA and CERAPP projects, and the findings of Cirino et al. following the Tox24 Challenge.

In conclusion, this research has not only provided new tools to support TH system disruption assessment by alternative methods but has also contributed to consolidating existing knowledge gaps and provided a clear perspective on how to advance the field. The successful application of the QSAR-AOP integrated approach provided a solid foundation for future research and regulatory applications. Further work could involve expanding this methodology to other MIEs for TH system disruption, as well as to other AOPs to assess a broader range of endocrine-disrupting mechanisms. Additionally, the robust and predictive models developed in this research, which align with the OECD guidelines, can be refined and validated with larger, more diverse datasets to further boost their already strong predictive power and regulatory acceptance. The integration of these models and other different NAMs within the AOP framework, in the context of integrated approaches for testing and assessment (IATA), should be explored to achieve a more holistic approach, building more comprehensive and robust assessment tools. To conclude, this work contributed to the broader goal of reducing reliance on animal testing and accelerating the identification of hazardous substances, thereby supporting more sustainable and safer chemicals.

List of Publications

Scientific articles

I. Evangelista, M., Chirico, N., and Papa, E., 2024. *In Silico* Models for the Screening of Human Transthyretin Disruptors. Journal of Hazardous Materials, 480, 136188, DOI: 10.1016/j.jhazmat.2024.136188.

II. Evangelista, M., Chirico, N., and Papa, E., 2025. New QSAR Models to Predict Human Transthyretin Disruption by Per- and Polyfluoroalkyl Substances (PFAS): Development and Application, Toxics, 13 (7), 590, DOI: 10.3390/toxics13070590. Selected as a Feature Paper in Toxics.

III. Evangelista, M., and Papa, E., 2025. A Review of Quantitative Structure–Activity Relationship (QSAR) Models to Predict Thyroid Hormone System Disruption by Chemical Substances, Toxics, 13 (9), 799, DOI: 10.3390/toxics13090799.

IV. Sgariboldi, A., Posté, E., Chirico, N., Sangion, A., Evangelista, M., Morosini, C., Re, A., Torretta V., and Papa, E., 2025. Global Assessment of Emerging Contaminant Removal in Wastewater Treatment Plants: *In Silico* Hazard Screening and Risk Evaluation", 13 (6), 1, DOI: 10.3390/toxics13010006.

V. Part of the research conducted at the European Chemicals Agency (ECHA) contributed to a new manuscript which is currently under revision.

Oral presentations at conferences

I. Evangelista, M., Chirico, N., Papa, E. State-of-the-art analysis of QSAR models for Thyroid Disruption-related endpoints. Joint Research Centre (JRC) Summer School, Ispra, Italy, 23 - 26 May 2023.

II. Evangelista, M., Chirico, N., Papa, E. *In silico* prediction of Thyroid disruption-related endpoints: state-of-the-art analysis of existing QSARs. 2nd Workshop SETAC Italian Language Branch, Rome, Italy, October 11th 2023.

III. Evangelista, M., Chirico, N., Papa, E. QSARs For the Prediction of the Human Transthyretin Binding Affinity of Per- And Polyfluoroalkyl Substances. 21st International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2025), Milan, Italy, 03 - 06 June 2025.

IV. Evangelista, M., Chirico, N., Papa, E. *In Silico* Prediction of the Human Transthyretin Binding Affinity. 3rd Workshop SETAC Italian Language Branch, Milan, Italy, 07 - 08 October 2024.

V. Postè, E., Chirico, N., Sgariboldi, A., Evangelista, M., Morosini, C., Re, A., Karaeva, A., Papa, E. Hazard evaluation and prioritization of emerging contaminants in the aquatic environment using *in silico* approaches". 1st International Conference on Circularity, Sustainability and Resilience in Water, Wastewater and Sludge Management (CSRW24), Varese, Italy, 11 - 13 February 2024. Coauthor of oral presentation by Sgariboldi, A.

VI. Sgariboldi, A., Posté, E., Chirico, N., Sangion, A., Evangelista, M., Morosini, C., Re, A., Torretta V., Papa, E. *In silico* screening of hazardous properties of Pharmaceuticals and Personal Care Products, and their associated potential risk after water treatment. 3rd Workshop SETAC Italian Language Branch, Milan, Italy, 07 - 08 October 2024. Coauthor of oral presentation by Sgariboldi, A.

VII. Sgariboldi, A., Posté, E., Chirico, N., Sangion, A., Evangelista, M., Morosini, C., Re, A., Torretta V., Papa, E. Evaluation of PPCPs In Wastewater Treatment Plants: Combining QSAR Modeling with Monitoring Data. 21st International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2025), Milan, Italy, 03 - 06 June 2025. Coauthor of oral presentation by Sgariboldi, A.

VIII. Chirico, N., Evangelista, M., Sgariboldi, A., Papa, E. QSAR-ME Profiler 2025, A New Software for Screening Target Chemicals and Metabolites Supported by Similarity Analysis and Domain Inspection. 21st International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2025), Milan, Italy, 03 - 06 June 2025. Coauthor of oral presentation by Papa, E.

IX. Sgariboldi, A., Paredi, F., Evangelista, M., Chirico, N., Papa, E. Integration of *in silico* methods into life cycle and sustainability assessments. 4th Workshop SETAC Italian Language Branch, Naples, Italy, 13 - 14 October 2025. Coauthor of oral presentation by Sgariboldi, A.

Poster presentation at conferences

I. Evangelista, M., Chirico, N., Papa, E. An overview of QSAR models for the prediction of thyroid disruption-related endpoints. 18th International

Conference on Chemistry and Environment (ICCE), Venice, Italy, 11 - 15 June 2023.

II. Evangelista, M., Chirico, N., Papa, E. State-of-the-art analysis of QSAR models for Thyroid Disruption-related endpoints. Joint Research Centre (JRC) Summer School, Ispra, Italy, 23 - 26 May 2023.

III. Evangelista, M., Chirico, N., Papa, E. *In Silico* Classification Model to Screen the Potential Thyroid Hormone System-Disrupting Activity of Per- and Polyfluoroalkyl Substances. SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025.

IV. Evangelista, M., Chirico, N., Papa, E. *In Silico* Tools for Endocrine Disruption Assessment: A Focus on Human Transthyretin Disruptors. SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025.

V. Evangelista, M., Chirico, N., Papa, E. QSAR Models to Predict the Human Transthyretin Binding Affinity of Compounds. 21st International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2025), Milan, Italy, 03 - 06 June 2025.

VI. Sgariboldi, A., Posté, E., Chirico, N., Sangion, A., Evangelista, M., Morosini, C., Re, A., Torretta, V., Papa, E. Quantification of the Potential Hazard and Risk of Pharmaceutical and Personal Care Products in Wastewater Treatment Plants, Combined with Hazard and Risk Screening Using *In Silico* Approaches". SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025.

VII. Chirico, N., Evangelista, M., Sgariboldi, A., Papa, E. QSAR-ME Profiler 2025: QSAR predictions, similarity analysis, domain inspection and metabolites profiling. SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025. Presented by Evangelista, M. also as a Poster Spotlight.

VIII. Talja, P., Evangelista, M., Nyman, A.-M., Furuno, S., Simanainen, U., Craenen, K., Sobanska, M. Aquatic Toxicity Profile of Neurotoxic Substances. SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025.

IX. Simanainen, U., Evangelista, M., Talja, P., Nyman, A.-M., Furuno, S., Craenen, K., Sobanska, M., Sobanski, T. *In Silico* Pre-Screening to Identify Neurotoxic Substances Which Are Considered Out of Applicability Domain for The Fish Embryo Acute Toxicity (FET) Test. SETAC Europe 35th Annual Meeting, Vienna, Austria, 11 – 15 May 2025.

X. Moncho, S., Evangelista, M., Simanainen, U., Talja, P., Nyman, A.-M., Furuno, S., Craenen, K., Sobanska, M., Sobanski, T. 21st International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences (QSAR2025), Milan, Italy, 03 - 06 June 2025.

XI. Chirico, N., Bertato, L., Evangelista, M., Papa, E. Computational tools for the development and application of QSARs for the estimation of multiple

endpoints. SETAC Europe 33nd Annual Meeting, Dublin, Ireland, 30 April – 4 May 2023.

Datasets

This section includes references to some unpublished data collected in the context of paper I presented in Chapter 2 (dataset A) and paper II presented in Chapter 4 (dataset B) of this thesis.

A. Evangelista, M., Papa, E. Human transthyretin (hTTR) binding affinities of chemical substances. https://doi.org/10.5281/zenodo.17278792.
B. Evangelista, M., Papa, E. Human transthyretin (hTTR) binding affinities of per- and polyfluoroalkyl substances (PFAS). https://doi.org/10.5281/zenodo.17279729.

Acknowledgments

My deepest gratitude goes to my supervisor, Professor Ester Papa. Thank you for the incredible opportunity, for welcoming me into your research, and for your strong belief in my potential from the very beginning. I am profoundly thankful for your constant availability, your guidance, and the support you provided, not just during the good times, but especially during the most challenging ones. Your mentorship has truly made a difference, and I will always be grateful.

I also want to thank Dr. Nicola Chirico for generously sharing his experience and providing valuable advice throughout this entire project. His input was a crucial resource for this research.

A special thank you to Arianna for sharing the ups and downs of the doctoral journey. Your constant presence and source of support made the demanding process feel much easier.

I am deeply grateful to the European Chemicals Agency (ECHA) for the fantastic opportunity they provided me. My sincere thanks go to Pia, Anna, and Ulla for their constant trust and support, which allowed me to have such an incredible professional experience. My deep gratitude also extends to everyone else who helped make my time at ECHA both stimulating and enjoyable.

I would like to sincerely thank Professor Davide Ballabio and Professor Olivier Taboureau for taking the time to carefully review this thesis. I truly appreciated their insightful comments and the attention and time they dedicated to evaluating this work.

I owe a profound debt of gratitude to my family. To my parents, Mauro and Ornella, and my sister, Giulia. Thank you for always supporting my choices and for your unwavering encouragement from the very beginning to the conclusion of this journey. Their unconditional love made all the difference.

To my Ariel, whose arrival illuminated my life with boundless joy and love.

Lastly, but by no means least, my heartfelt thanks go to those friends who understood me and stood by me throughout this journey.