Joint Research Programme
BTO 2024.016 | January 2024

Evaluation of innovative methods to assess neuroactive substances in water

Joint Research Programme

KWR

Bridging Science to Practice

Evaluation of innovative methods to assess neuroactive substances in water

BTO 2024.016 | January 2024

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045/354

Project manager
Patrick Bäuerlein, PhD

Client
BTO - Verkennend Onderzoek (VO)

Author(s)
Sanah Majid, MSc., ERT

Renske Hoondert, PhD

Daniel Duarte, PhD

Astrid Reus, MSc.

Quality Assurance
Milou Dingemans, PhD, ERT

Tessa Pronk, PhD

Sent to
This report is distributed to BTO-participants.

A year after publication it is public.

Keywords: neurotoxicity, *in vivo, in silico*

Year of publishing
2024

More information
Sanah Majid, MSc. ERT
T +31 30 606 9716
E Sanah-Majid.Shaikh@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

Managementsamenvatting

De combinatie van een QSAR-methode en effectmeting in een platwormmodel helpt om potentiële risico's van neuroactieve stoffen in water te beoordelen

Auteurs: Sanah Majid, Renske Hoondert, Daniel Duarte, Astrid Reus

Drinkwaterbronnen kunnen verontreinigd zijn met chemische stoffen die schadelijke effecten kunnen hebben op het zenuwstelsel. Onvoldoende kennis van de toxiciteit en de mechanismen die ten grondslag liggen aan de neurotoxiciteit van deze stoffen maakt het inschatten van de risico's moeilijk. Om deze kennisleemte op te vullen, is een *in silico* methode onderzocht gebaseerd op de Quantitative Structure-Activity Relationship (QSAR)-benadering en ongewerveld diermodel (planarian = platworm) om de neurotoxiciteit te voorspellen van chemische stoffen die relevant zijn voor water. Deze combinatie van innovatieve methoden maakt het mogelijk de waterkwaliteit te beoordelen door theoretische voorspellingen effectief te koppelen aan werkelijke neurotoxische effecten die worden waargenomen in een platwormmodel. Verder onderzoek zal bijdragen aan de validatie van deze combinatie die theoretische voorspellingen en empirische bepalingen combineert.

Belang: Risico's van neuroactieve stoffen in drinkwaterbronnen zijn nog onduidelijk

Er komen steeds meer chemische stoffen uit verschillende bronnen in watersystemen terecht. Het wordt dus ook steeds belangrijker om te begrijpen wat de mogelijke gezondheidseffecten van deze chemische verontreinigingen zijn. Vooral om de neurotoxiciteit van dergelijke verontreinigingen te bepalen zijn nog weinig testen of effectmetingen beschikbaar. Een uitgebreidere en gerichtere beoordeling van de neurotoxiciteit is van cruciaal belang om potentiële bedreigingen door neuroactieve stoffen te identificeren en aan te pakken, als ondersteuning voor de bewaking, de beoordeling en het beheer van potentiële risico's door neuroactieve stoffen in drinkwaterbronnen.

Aanpak: Combinatie van *in silico*-instrumenten en ongewervelde neurotoxiciteitsbioassay

In deze studie werden QSAR-modellen (Quantitative Structure-Activity Relationship) afgeleid voor de voorspelling van neurologische ontwikkelings-toxiciteit van chemische stoffen. Deze werden gebaseerd op neurotoxiciteitsgegevens uit de ToxCast database, waarbij functionele groepen (structurele elementen van chemische stoffen die geassocieerd worden met toxiciteit) werden gebruikt als verklarende variabelen in meervoudige lineaire regressiemodellen. De modelprestaties, gevoeligheid (gerelateerd aan het aantal correct voorspelde ware positieven) en specificiteit (gerelateerd aan het aantal correct voorspelde ware negatieven) werden geëvalueerd. Verder werd er een diepgaand literatuuronderzoek uitgevoerd om inzicht te krijgen in de huidige ontwikkelingen op het gebied van neurotoxiciteitstesten en de geschiktheid van op platworm gebaseerde bioassays (planarian bioassays) voor het evalueren van de neurotoxiciteit van waterrelevante verontreinigingen.

Resultaten: het combineren van *in silico* tools met neurotoxiciteit-bioassays voor monitoring

Het literatuuronderzoek heeft inzicht opgeleverd in ontwikkelingen in de regelgeving voor de beoordeling van neurotoxiciteit. De integratie van *in silico* (QSAR) en *in vivo* methoden voor het bepalen

van neurotoxiciteit met behulp van een alternatief diermodel (platworm of *planarians*) voor de Nederlandse en Vlaamse watersector biedt potentie voor bewaking van de chemische waterkwaliteit. Aangezien neurotoxiciteit een complex toxicologisch eindpunt is, is een nauwkeurige schatting met een enkele testmethode een uitdaging. Een waardevolle aanpak is daarom een stapsgewijze strategie die *in silico* -en alternatieve *in vivo*-modellen (zoals planaria) combineert. Standaardiseren van deze strategie kan de nauwkeurigheid bij het screenen van neuroactieve stoffen in drinkwaterbronnen vergroten.

Toepassing: Naar validatie en diepere inzichten door verder onderzoek

De beschreven benaderingen integreren QSAR en een planarian bioassay tot een potentieel alomvattende en stapsgewijze benadering om de potentiële neurotoxiciteit te beoordelen van chemicaliën die relevant zijn voor water. Deze aanpak maakt een effectieve prioritering van verontreinigende stoffen mogelijk. Om een transparante, consistente en betrouwbare beoordeling van neurotoxiciteit op basis van chemische structuren te realiseren, zijn echter gestandaardiseerde methoden nodig. Bovendien moeten de beperkingen bij het voorspellen van complexe biologische reacties worden erkend. Validatie achteraf met behulp van het platwormmodel maakt vergelijking mogelijk tussen QSAR-voorspellingen en de werkelijke neurotoxische effecten en doet recht aan de inherente complexiteit van biologische systemen. Deze strategische combinatie biedt mogelijkheden tot standaardisatie en voor gebruik op grote schaal om de water-kwaliteitsmonitoring te ondersteunen en zo de waterkwaliteit en de volksgezondheid te beschermen.

Rapport

Dit onderzoek is beschreven in het rapport *The evaluation of innovative methods to assess neuroactive substances in water* (BTO 2024.016).

Contents

More Information
Sanah Majid, MSc, ERT
T +31 30 606 9716
E Sanah-Majid.Shaikh@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

Summary

Neurotoxicity screening is a rapidly developing field in which the adverse effects of substances are assessed to identify those chemicals that have the potential to affect the nervous system. Scientific evidence shows that even mild impairments in the nervous system can lead to transient or persistent effects on cognitive and mental health, and each of these can be inflicted via direct or indirect exposure. Based on known modes of action (MoA) of organic contaminants detected in freshwater monitoring studies, neuroactive compounds represent the largest category, comprising pharmaceuticals, pesticides and illicit drugs, i.e., almost 30% of all detected chemicals (Busch *et al.*, 2016, Legradi *et al.*, 2018). This emphasises the need for tools for rapid exposure and effect assessments to prevent possible harm to public health. Exposure to neuroactive compounds occurs via different routes. However, a good understanding of the links between exposure to these substances and the mechanism by which these affect the nervous system remains largely unknown. Therefore, there is a need for consistent guidance on the development of relevant test methods and the reliable evaluation of data on neuroactive substances for assessing their potential risk. Limited testing capacity has led regulators to encourage the development and use of more informative New Approach Methods, including *in vitro, in silico* and alternative *in vivo* approaches to predict the neurotoxicity or developmental neurotoxicity (DNT) of chemicals.

Our earlier BTO research emphasised the need to stay attentive to neuroactive substances in the water system, and monitor the developments in bioassays for neuroactive substances (BTO 2020.035; Reus et al., 2020). Available models and techniques suitable for neurotoxicity testing, and their potential for water quality monitoring were listed. It was concluded that more research is needed to define which of the models and techniques are most appropriate and applicable for neurotoxicity testing and what steps are needed to implement them. In this follow-up project the application of an *in silico* method (using Quantitative Structure-Activity Relationships models: QSARs) and an *in vivo* bioassay (using planarians) for neurotoxicity testing within water quality assessment are presented. The report comprises of two sections. In the first section, an *in silico* method is described and the prediction of the neurotoxicity (using *in vitro* assay data from the ToxCast database) of compounds is evaluated by determining the robustness and accuracy of the predictions for compounds outside the training dataset (Dix et al., 2007). The *in vivo* method is described in the second section and is based on the guidelines described by Hans et al., (2012) and the selection criteria for the implementation and use of bioassays described in the DEMEAU report (Schriks et al., 2012).

This report shows the applicability of the proposed methods, offering a comprehensive analysis of their cost-effectiveness, skills, implementation for water samples and also provides a description of robustness and throughput capacity of their performance. Both methods are complementary and applicable for neurotoxicological screening of substances including those for which there is insufficient data as well as emerging chemicals and transformation products. The report envisions pilot studies as a follow-up project, in which a test set comprising water-relevant substances and water samples will be studied. This initiative aims not only to validate and refine the methods but also allows testing of additional endpoints such as genotoxicity, carcinogenicity and developmental toxicity of water relevant substances. The inclusion of these additional endpoints will broaden the scope, providing insights to inform water quality managers and prompt necessary actions such as improving treatment processes. Furthermore, the integration of QSAR in combination with a simplified *in vivo* method will increase the weight of evidence and provide interpretable results as to whether a chemical has the potential to cause neurotoxicity. Thus ensuring more informed actions in safeguarding water quality.

1 Introduction

1.1 Neurotoxicity

Neurotoxicity is defined as an adverse change in the structure or function of the nervous system [central nervous system (CNS) and/or peripheral nervous system (PNS)] following exposure to a chemical or a physical agent (WHO/IPCS 2001). Neurotoxic substances may be naturally occurring (neurotoxins), e.g., Botulinum toxin produced by the bacterium *Clostridium botulinum* and related species, and/or synthetically produced (neurotoxicants) e.g., certain food additives, pesticides, pharmaceutical drugs, illicit drugs, and industrial chemicals (Bilge, 2022). Neurotoxicants can alter the activity of the nervous system which can lead to neurodegeneration or death of neurons (Hubbs-Tait et al., 2005). Neurons are a key component of nervous tissue responsible for transmitting and processing electrochemical signals in both the central and the peripheral nervous system. Together with glial cells neurons are a primary component of the nervous system. Glial cells which are non-neuronal cells play a crucial role in providing structural and metabolic support to the nervous system (Moore et al., 2005). The structural organisation of a neuron is depicted in Figure 1. A neuron consists of a cell body containing the nucleus dendrites, and an axon. Often multiple neurons form clusters, called nerve nodes. Dendrites are thin, branched structures that extend from the cell plasma of the cell body. They form a surface that receives stimuli and transmits electrical impulses to the cell body. Neurons communicate with other cells via synapses (specialised connections) that use neurotransmitters (signalling molecules) to pass the electric signal from the presynaptic neuron to the target cell (citations). When information is received from an external or internal stimulus, it is transmitted down the axon as a nerve impulse. Eventually, this information ends up at the axon terminal, where it is chemically transmitted to another neuron.

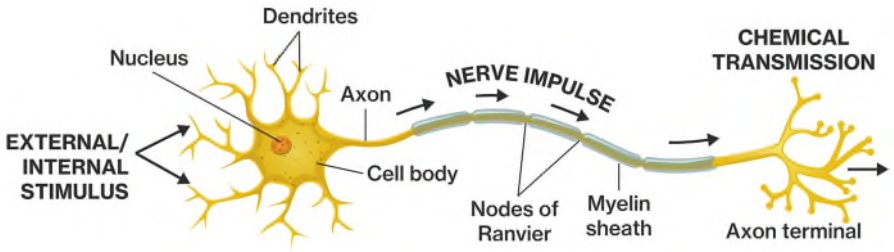

Figure 1: Structure of a Neuron. *Image source: https://www.coursehero.com/sg/anatomy-and-physiology/neuron-structure/*

Some neurotoxic substances may act directly on neurons, while others may act indirectly by affecting metabolic processes on which the neurological system depends (Wright et al., 2009; Bilge, 2022). These effects can be observed at different levels of biological organization, from neurochemical alterations (e.g., alteration of ion flow across cell membranes, or prevention neurotransmitter release from the nerve terminals), anatomical alterations (alterations of the cell body, axon, or myelin sheath), physiological alterations (alterations in the thresholds for neuronal activation or reduction in the rate of neurotransmission), to behavioral alterations (e.g., alterations in motor or cognitive functions) (USEPA, 1998; Fitzgerald et al., 2020; Robertson, 2022; Kamińska and Cudnoch-Jędrzejewska, 2023). Furthermore, the effects may be transient (i.e., reversible) or persistent (i.e., permanent) (Bilge, 2022). In addition, the effect of some neurotoxic substances may be primary (do not require metabolism prior to interacting with target sites) or secondary (require metabolism prior to interacting with target sites) (WHO/IPCS, 2001; Bilge, 2022).

Neurotoxicity can occur at any stage of life and the effects can change with age (NCBI 1992; Xia et al., 2021). Scientific evidence suggests that the developing nervous system is more susceptible, and/or differentially susceptible, to toxic exposures than the adult nervous system (Giordano and Costa, 2012; Crofton et la., 2022). The attributed cause of this phenomenon is the complexity of developing brain and the developmental window at the time of exposure (Rice and Barone et al., 2000; Crofton et al., 2022). Disturbances during the developmental window can lead to adverse effects on the nervous system. The adverse effects can be a result of the alterations in both kinetic and dynamic processes (Crofton et al., 2022). The kinetic processes (absorption, distribution, metabolism and excretion) comprise the blood-brain-barrier function and the metabolic capacities of different cell types, which result in differences in the concentrations of toxicants at the active sites (OECD/EFSA, 2016). The dynamic processes pertain to chemical interactions involving biochemical, molecular and physiological processes such as the regulation of cell proliferation, survival and differentiation (Nguyen et al., 2001; Giordano and Costa, 2012). These processes can be susceptible to chemical stressors, some of which may have specific effect on the developing brain. Furthermore, the development of nervous system takes longer than the development of other organs (Rodier, 1994). This extended duration increases its susceptibility to chemical insults over a prolonged period

1.2 Developments in Neurotoxicity Assessment

Interest in neurotoxicological research has increased greatly in recent years because of the susceptibility of the nervous system to chemical exposure. The potential effects of neurotoxicity are very diverse and reflect the important role that the nervous system plays in regulating vital bodily functions. Impairment of these neurological processes (perception, evaluation, and response to the external environment) by overexposure to chemicals can produce a variety of abnormal neurological responses, some of which may be life-threatening, while others may result in short-term, permanent, or progressive neurological changes, including psychiatric disorders. Of particular concern is the possibility that exposure to certain neurotoxic substances, such as lead, may irreversibly impair the normal development and function of the human brain (Schneider et al., 2023).

In the recent years, there has been a consistent surge in the use and discharge of chemicals into the environment, which has led to increased scientific and regulatory attention towards the potential risks to human health caused by the exposure to neuroactive substances (Landrigan et al., 2012; Gadaleta et al., 2022). Risk assessment guidelines and regulations for chemicals, such as the European Union's REACH regulation (European Commission, 2006) and the United States' Toxic Substances Control Act (TSCA) (McPartland et al., 2022), prohibit or restrict harmful or very persistent chemicals as to protect humans and other organisms. While certain toxicological endpoints, such as endocrine disruption, have been recognised in hazard and risk assessment and included in restrictive guidelines, neurotoxicity remains a less focused area (Braun and Escher, 2023). Despite over 80,000 chemicals currently registered for commercial use by the US Environmental Protection Agency (USEPA), only a small fraction (approximately 200) have been studied for their potential developmental neurotoxic properties according to the established guidelines (Grandjean et al., 2006; Rock and Patisaul 2018; Morris et al., 2022). Furthermore, limited knowledge exists about which of these chemicals elicit developmental neurotoxicity, and systematic screening for such properties before commercial adoption remains (Rock and Patisaul 2018).

Traditionally, industries, including the chemical, pharmaceutical, and food sectors have heavily relied on *in vivo* research involving rodents to assess potential neurotoxicity during development (Reus et al., 2020). As of now, alternative toxicity testing methods are not legally accepted. There are three *in vivo* tests that are legally accepted: OECD Test Guideline 424 (Neurotoxicity Study in Rodents), OECD Test Guideline 426 (Developmental Neurotoxicity Study) and OECD Test Guideline 443 (Extended One-Generation Reproductive Toxicity Study). While *in vivo* studies continue to serve as the basis for toxicological evaluation across different endpoints, recent regulatory shifts

increasingly emphasize the need for the incorporation of alternative testing methods both for neurotoxicity and *in vivo* studies in general. The European Food Safety Authority (EFSA), the European Chemicals Agency (ECHA), and the United States Environmental Protection Agency (US EPA) currently encourage the development and use of alternative testing methods. This includes the New Approach Methodologies (NAMs) such as *in vitro* and *in silico* approaches and the use of alternative *in vivo* models (invertebrates or vertebrate embryos) to predict the adult neurotoxicity (NT) or developmental neurotoxicity (DNT) of chemicals (Paparella et al., 2020; EFSA; 2021; Crofton et al., 2022; EFSA, 2022). Furthermore, the Tox21 consortium (Toxicology in the 21st Century, accessible via: https://tox21.gov/overview/) initiative aims to establish a new area of toxicological testing for chemicals for which little or no testing had been done. Tox21 aims to revolutionise toxicological testing by combining alternative non-animal approaches including *in silico* with alternative *in vivo* models (invertebrates and lower vertebrates). This approach facilitates improvement of screening throughput as they can be automated at lower cost and are free from ethical issues (Hagstrom et al., 2016). In addition, many genes and important metabolic pathways are conserved between lower level organisms and humans, so that their use for molecular studies enables mechanistic insights into the toxicity of compounds (Lein et al., 2005).

2 . Drinking water chemicals and risk of neurotoxicity

The extensive use of chemicals has resulted in a continuous influx of new chemicals into freshwater and groundwater supplies, thereby exerting pressure over or posing a threat to the drinking water quality. These substances can exist as parent compound or transformation product, for example as an active metabolite. Often, toxicity of many of these substances is poorly understood as well as their underlying toxicity mechanisms. Monitoring programs to evaluate the toxicity potential of surface water and treated wastewater may not include many of these chemicals due to their poor characterization, including their neurotoxic potential. Studies have shown that most compounds detected in European freshwater systems have the potential to be neurotoxic and may interact with the nervous system via different molecular mechanisms and targets (Zhou et al. 2019; Kaisarevic and colleagues, 2021). The widespread use and discharge of anthropogenic substances has resulted in their presence in Dutch and Flemish water bodies (Kools et al. 2019). Given that a significant proportion of Dutch (approximately 40%) and Flemish (approximately 50%) drinking water is derived from the (infiltrated) surface water, understanding the potential adverse effects of neuroactive substances in the aquatic environment becomes crucial. Consequently, neurotoxicity emerges as a relevant toxicological endpoint for chemical water quality. Measurements of relevant effects can be added to a water quality testing strategy, such as the 'Key factor toxicity' bioassay set (SF toxiciteit, 2023), that was based on smart Integrated Monitoring (SIMONI), a bioanalytical strategy for water quality assessment for determining surface water quality (Van der Oost et al., 2017). Despite the increasing prevalence of neuroactive substances in the aquatic environment, tests to evaluate neurotoxicity potential in water quality assessment have been less applied. This could be attributed to several factors, including the conventional focus on acute effects on human health, whereas neurotoxic effects manifest over a longer period, making them less apparent in short-term assessments (Giordano et la., 2012; Spencer and Lein, 2024). In addition, neurotoxicity endpoints are not as well-established within regulatory frameworks, resulting in a lack of standardisation in assessment (Crofton et al., 2022). Furthermore, the development of standardised methods for assessing neurotoxicity in water systems are still evolving. Additionally, the availability of cost effective, reliable and efficient neurotoxicity assays for routine monitoring remains a challenge. Other possible factors contributing to limited consideration of evaluating neurotoxicity potential in water relevant substances may be linked to limited data on new compounds, lack of public awareness on potential long-term neurotoxic effects of water relevant substances, costs and resource constraints. Addressing these challenges requires concerted efforts from regulators, water sector and toxicologists to develop standardised methods, raise awareness and allocate resources for the inclusion of neurotoxicity assessment in routine water quality monitoring.

A previous BTO report highlighted the need to monitor developments in bioassays for neuroactive substances and listed the available models and techniques suitable for neurotoxicity testing, in the context of chemical characterization and their potential for water quality monitoring (BTO 2020.035; Reus et al., 2020). It was concluded that more research was needed to define which of the candidate models and techniques are most suitable and what steps are needed to implement them in practise. This follow-up project will evaluate the usability and value of innovative methods for neurotoxicity testing in water quality assessment. As our understanding of harmful effects on nervous system advances, it is likely that there will be an increasing emphasis on incorporating these assessments into a broader water quality management strategies. In the realm of neurotoxicity assessment within the context of water quality, we have identified two approaches particularly relevant; an *in silico* prediction approach Quantitative Structure-Activity Relationship (QSAR) models and an *in vivo* approach planarian flatworms. Application of *in-silico* approach using QSAR and the *in vivo* bioassay using planarian flatworm for neurotoxicity testing, with special focus on DNT within water quality assessment will facilitate risk-based monitoring strategies at water companies. The selection of the *in vivo* model is based on the guidelines described by Hans et al., (2012) and the selection criteria for *in vitro* bioassays implementation and use described in the DEMEAU report (Schriks et al., 2012).

3. *In silico* method for neurotoxicity testing

In silico methods (i.e. computer models) to predict toxicity or activity have been used widely in pharmaceutical sciences and drug development (Grummt et al., 2022). However, this approach has not been completely established in neurotoxicology, and QSAR (Quantitative Structure-Activity Relationship) models for neurotoxicity are rare (Jiang et al., 2020). QSAR models that have been developed in the 1990s and 2000s to predict neurotoxicity has had only limited success or were only restricted to certain specific endpoints (Grummt et al., 2022). Since a strong relationship between chemical structure and neurotoxicological activity has only been recently established and abundant experimental data for neurotoxicity are lacking, many commercial and public tools to perform read-across (e.g. OECD QSAR Toolbox and VEGA Hub) lack modules for neurotoxicological assessment.

In 2020, (Jiang et al., 2020) collected systemic neurotoxicity data (LD_{50}s – lethal dose at which 50% of the organisms is affected - in mice and rats) for 422 compounds to construct QSAR models using eight different machine learning models. Molecular descriptors taken from the PyBioMed software were taken as explanatory variables in these models. In the study by Jiang et al., sample size in the model construction was restricted by the lack of sufficient *in vivo* neurodevelopment LD_{50} data. Sample size may be increased by basing the QSAR model development on specific *in vitro* endpoints taken from the ToxCast database. In the presented study, neurotoxicity or neurodevelopmental toxicity *in vitro* assays from the ToxCast database were included in multiple linear regression models as response variables. This mechanistic neurotoxicity data is way more relevant for water pollutants than lethal doses for systemic neurotoxicity. Structural molecular parameters (i.e. the functional groups and topological descriptors) were taken as explanatory variables and were taken from the *rcdk* package (R – Chemistry Development Kit, a collection of modular Java libraries for processing chemical information) in R, as the PyBioMed software was unavailable and dependencies in the accompanying Python scripts were failing. When including these functional groups in a linear regression model as dummy explanatory variables, R^2s (based on training datasets) and Q^2s (based on test datasets) as high as 80% can be obtained, although in some cases, the model outcomes may be overfitted (when the R^2 is very high and the Q^2 is below zero, which is the case in many of the models including interaction terms). Additionally, subdividing the toxicity data in categories based on thresholds (lower 25th percentile of toxicity data (AC_{50}; the concentration at which 50% of the activity is observed) – high toxicity; mid-50th percentile- medium toxicity; higher 75th percentile of the data – low toxicity) leads to fairly good predictions, with models yielding high sensitivity, specificity and accuracy values. The sensitivity, specificity and accuracy are the abilities of the tests to recognize toxic substances (true positives), recognize non-toxic substances (true negatives) and the general predictive ability of the model(true positives + true negatives), respectively. The highest predictive power (in terms of R^2, Q^2, sensitivity, specificity and accuracy) could

be obtained for a model derived based on *in vitro* assays for synaptogenesis (See Appendix I.II). This endpoint has been suggested as the most sensitive developmental neurotoxicity endpoint to chemical mixture-induced effects (Pistollato et al., 2020). In the present study, we focused on predicting activities in *in vitro* assays rather than biological processes or systemic toxicity in complete organisms. Despite their increasing usage, the application of the AOP framework – like was done in (Gadaleta et al., 2022) - in computational toxicology is hampered by multiple challenges. Future research may benefit from in depth investigations to disentangle these processes in order to ultimately predict systemic toxicity in humans.

3.1 Methods

3.1.1 Data acquisition

To advance the *in silico* prediction of compounds with neurotoxic activity, in the present study neurotoxicity data were taken from the ToxCast database (Dix et al., 2007). This database consists of 21 databases, encompassing over 3.7 million experimental toxicity data records. In ToxCast, the active concentration at which 50% of the effect is observed (AC_{50} in µM) is calculated using experimental dose response series for a wide range of *in vitro* bioassays and three statistical model types (constant (two-parameter) model, Hill (three parameter) S-model, and gain-loss model, which is the product of two Hill models). Log AC_{50} values for the best predictive model are calculated. In ToxCast, concentration-response series only get an active hit call (and high quality rating) when they meet criteria, based on the number of Hill model 'wins' and sample size. More details on the method are presented in BTO report 2023.060.

ToxCast bioassay test results related to neurotoxicity were selected, based on bioassay intended target family (neurodevelopment) and subfamily (neural network function (activity, bursting activity, network connectivity), synaptogenesis/neurite maturation, and neurite outgrowth). Sub setting based on *in vitro* neurodevelopment assays resulted in a dataset with 3315 data rows, relating to 239 unique chemicals. These data were combined with data on several types of molecular descriptors:

1. Information on functional groups (nested from US EPA), taken from the QSAR Toolbox, resulting in a complete dataset including presence/absence data for 155 individual functional groups and 222 unique compounds.
2. These were combined with topological descriptors based on SMILES, using the *rcdk* package in R rcdk package - RDocumentation. SMILES were taken from EPI Suite, based on CAS number. However, as these sometimes resulted in invalid SMILES codes, additional SMILES were taken using the *webchem* package in R, using the CACTUS model webchem: Chemical Information from the Web (r-project.org). Topological descriptors in the *rcdk* package included in the present study are molecular weight, number of atoms, number of bonds, TPSA (topological polar surface area), XlogP, AlogP, charge, number and kind of atoms (i.e., sulfur, fluor, oxygen, nitrogen, and carbon). *XlogP* and *AlogP* both reflect the octanol/water partition coefficient (logP), calculated by summing the contributions from component atoms and correction factors in two separate atom-additive models (Ghose et al., 1998; Wang et al., 1997).

3.1.2 Model design

Neurodevelopmental toxicity was predicted based on quantitative structure-activity relationships (QSARs). The QSARs were derived by fitting the following conceptual models (multiple linear regression models) to the formatted data (excluding and including interaction terms to account for plausible synergism between functional groups and topological features and toxicity):

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n \quad [1]$$

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \cdots + \beta_o x_p x_q \quad [2]$$

in which β_1 to β_n represent the regression coefficients associated with the 1st to nth X_1 to X_n chemical property (either a physicochemical descriptor or a functional group), and y represents the toxicity endpoint (log_{10}-transformed AC_{50}).

3.1.3 Model performance evaluation

After combining the experimental neurotoxicity data with the topological descriptors and functional groups, 70% of the data were used as the training dataset and the remaining 30% served as the test dataset. The data was randomly allocated to these two lists, using the R *sample()* function. The models were evaluated using coefficient of determination (R^2) and cross-validated redundancy (Q^2) as validation metrics, for the model's ability to predict neurotoxicity of compounds in the test dataset (see Equation 3 and 4):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad [3]$$

where the R^2 is calculated as 1 – residual sum of squares (RSS) and the total sum of squares (TSS), y_i is the observed AC_{50} for compound i, $\hat{y}_i$ is the predicted AC_{50} for compound i, and $\bar{y}$ is the average AC_{50} in the training set. The R^2 statistics explains the variance in the response variable explained by the explanatory variable(s). Over the years, there has been ample discussion on the R^2 threshold above which a model can be considered a good predictive model. In the present study, R^2 values of 0.75, 0.50, or 0.25 for response variables will be described as substantial, moderate or weak, respectively (Hair et al., 2013; Sarstedt et al., 2021). The predictive power of the model was evaluated by calculating the Q^2 for the test dataset:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_{ext} - \hat{y}_{ext})^2}{\sum_{i=1}^{n}(y_{ext} - \bar{y}_{ext})^2} \quad [4]$$

where the Q^2 is calculated as 1 – residual sum of squares (RSS) and the total sum of squares (TSS), y_{ext} is the observed AC_{50} for compound i in the test set, $\hat{y}_{ext}$ is the predicted AC_{50} for compound I in the test set, and $\bar{y}_{ext}$ is the average AC_{50} in the test set. The Q^2 statistic reflects predictive relevance, and measures whether a model has predictive relevance or not. Q^2 values above zero indicate that your values are well reconstructed and that the model has predictive relevance. However, this does not say anything about the quality of the prediction, only that the model predicts better than taking the average of the observed values (Rigdon, 2014).

Because these evaluation criteria are based on nominal response variables (the substance is toxic or not), the continuous AC_{50} values per individual *in vitro* toxicity test have been converted into toxicity classes (low toxicity or high toxicity) based on the distribution of AC_{50} values (per toxicity test), whether the AC_{50} falls in the lowest 25% (high toxicity) or the highest 75% (low toxicity). The *sensitivity*, *specificity*, *accuracy, negative predictive value* (NPV) and *positive predictive value* (PPV) of the models to predict the correct bioassay response (high, low) were calculated based on these toxicity classes. Sensitivity (Equation 5) is the ability of the test to recognize toxic substances (true positives), specificity (Equation 6) is the ability of the model to recognize non-toxic substances (true negatives) and accuracy is the general predictive ability (true positives + true negatives; Equation 7). Whereas the sensitivity and specificity give information on the ability of the QSAR to correctly predict the known individual substances activity in the bioassay, the negative predictive value (NPV; the number of true negatives compared to the overall number of negative values; Equation 8) and positive predictive value (PPV; the number of true positives compared to all positive values; Equation 9) represent the proportion of a correct prediction of the bioassay activity with the QSAR, given the

QSAR result. This depends on the portion of substances with a respectively negative or positive response in a category.

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} \quad [5]$$

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \quad [6]$$

$$Accuracy = \frac{True\ positives + True\ negatives}{True\ positives + True\ negatives + False\ positives + False\ negatives} \quad [7]$$

$$NPV = \frac{True\ negatives}{True\ negatives + False\ negatives} \quad [8]$$

$$PPV = \frac{True\ positives}{True\ positives + False\ positives} \quad [9]$$

3.2 Results and Discussion

The predictive power of the multiple linear regression models when calculating AC_{50}s using models **per individual *in vitro* assay endpoint pertaining to neurodevelopment** was evaluated by comparing all predicted AC_{50}s (for all models per individual assay endpoint) with observed (experimental) values (Figure 2). Here, we see that 89.74% of all predicted data points fall within a factor of five of the experimental AC_{50}s. 0.11% of the data points overfitted (û-u=0; the predicted data is the same as the observed data), 4.62% of the data points underestimated (u/5 > û; the observed values were more than five times higher than the predicted data), 5.63% of the data points overestimated (u*5 < 5; the observed values were more than five times lower than the predicted data). However, there are significant differences in predictive power between the models including interaction terms and models excluding interaction among the explanatory variables, and between the training and test dataset. Overall, 79.5% of the variation in AC_{50} values in the training dataset is explained by the linear regression model (R^2), when excluding interaction terms between the explanatory variables. However, when the model is applied to the test data set, only 40.8% of the variation is explained by the model (Q^2). However, the explained variance in the test dataset is even lower when looking at the results for the linear regression model including interaction terms (Figure 2 – red text).

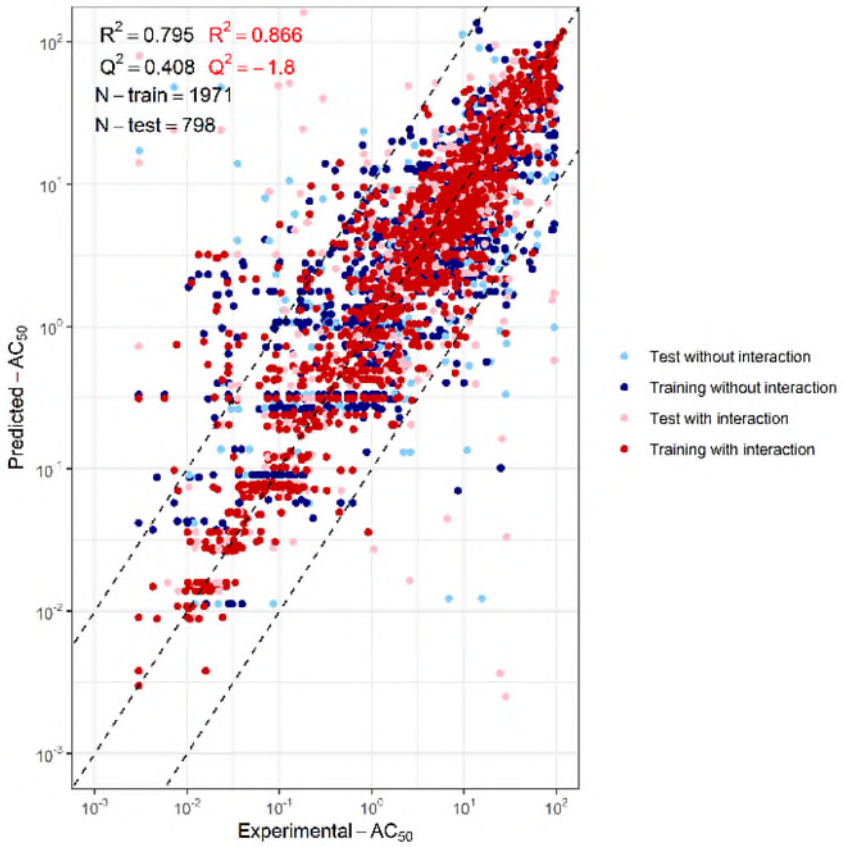

Figure 2. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus observed (experimental) neurotoxicity, based on functional groups and topological parameters. The R^2 (coefficient of determination for the training dataset) and Q^2 (prediction coefficient for the test dataset – outside the test dataset) are shown, when excluding (black text; blue dots) and including (red text; red dots) interaction terms.*

After grouping *in vitro* assays based on five subcategories of *in vitro* assay endpoints (neurite outgrowth (initiation), synaptogenesis and neurite maturation, neural network connectivity, network activity and network bursting activity), multiple linear regression analysis was based on structural fragments/functional groups as explanatory variables. Figure 3 describes to what extent the predicted AC_{50} values are within a factor of five of the observed AC_{50} values ('between') and whether the AC_{50} values are underestimated or overestimated compared to the observed values. The majority of the AC_{50} data in both the test and the training dataset could be well predicted by the derived models. In all cases (subcategory, test-training dataset, including and excluding interaction terms) over 75% of all predicted AC_{50} values fell within a factor of five from the observed (experimental) values. In general AC_{50} values tend to be underestimated (so more toxic than they really are) rather than overestimated. Overfitting of the models occurred only in the case of the models including interaction terms. While including interaction terms in general did improve the predictive power of the model for the training dataset (R^2), the predictive power of the model applied to the test dataset was generally lower (Q^2). This latter statistic is more important because the purpose of the QSAR is to predict neurotoxicity related-bioassay activity for previously untested, unknown chemicals. Overall, the highest model performance could be established for *in vitro* assay endpoints related to synaptogenesis and neurite maturation. For this category of endpoints also the smallest difference in R^2 and Q^2 could be observed (86.8% vs. 83.4%, respectively), implying that the model predicts equally well for data on which the model was based as well as chemicals outside the training dataset. More details with respect to the results for each category can be found in Appendix I.I. – I.IV.

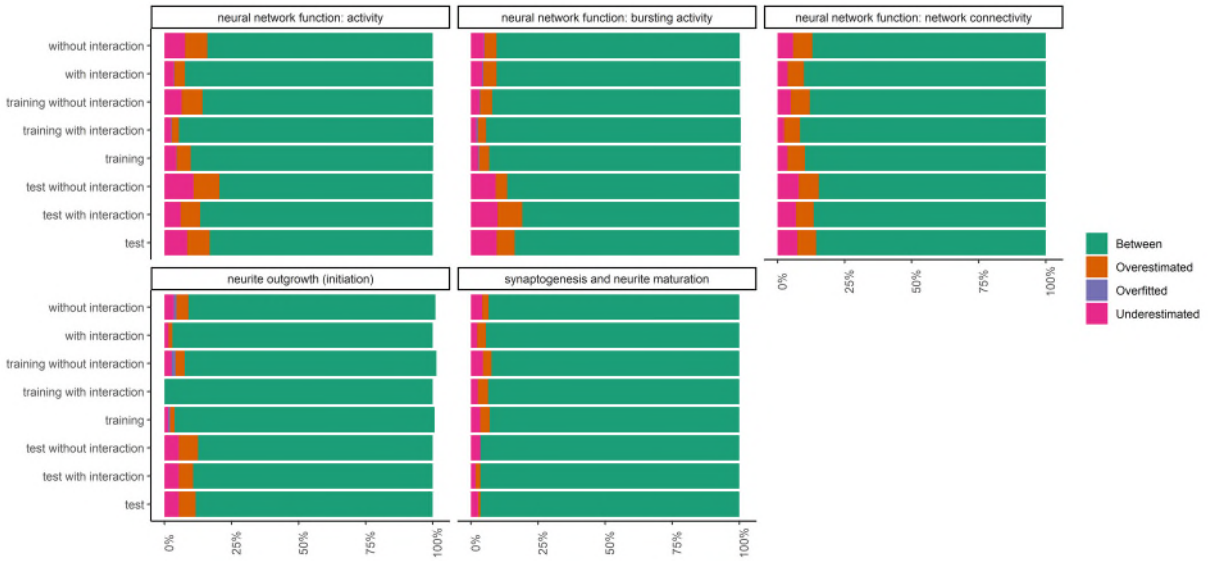

Figure 3. *Predictions of neurotoxicity data (percentages of predicted toxicity data that are within or outside a factor of five of true toxicity data), when in vitro assays are clustered based intended target family (neurodevelopment) and on intended target subfamilies (neural network function: activity, neural network function: bursting activity, neural network function: network connectivity, neurite outgrowth, synaptogenesis and neurite maturation).*

The current study was not the first to focus on the prediction of neurodevelopmental toxicity based on structural parameters of compounds. Gadaleta et al. (2022) performed three types of machine learning techniques (Random Forest, k-nearest neighbour and a neural network (MLP-NNET)). This study focused on using relatively non-complicated linear regression models - and two explanatory variable types (structural fingerprints and molecular descriptors) to predict neurodevelopmental toxicity, expressed as individual molecular initiating events (MIEs) in adverse outcome pathways (AOPs). This exercise resulted in good predictions, leading to accuracy values of at least 84%, and sensitivities and specificities of over 74% and 93%, respectively, depending on the individual MIEs. These values are in line with accuracy, specificity and sensitivity values calculated in the present study (chapter 3.2.1.). However, the dataset used in the study by Gadeleta et al. was slightly smaller than the dataset used in the present study (69 compounds compared to 203 compounds in the present study), although there is some overlap in compounds, as they used a subset of the ToxCast database. Due to the diverse assay technologies and study designs available in the ToxCast database, a highly generalized and robust (arithmetic median and median of absolute deviation vs arithmetic mean and standard deviation) set of calculations were performed (U.S. Environmental Protection Agency (EPA), 2014). However, the ToxCast program has acknowledged that false positive and false negative hit calls are possible using the automated methods, and have thus added a processing step to assign “flags” or warnings to the data (Ryan and Becker, 2017). Ryan and Becker (2017) describe possible flags in the ToxCast dataset that may be considered when analyzing a list of possible results. However, they also note that their assignment is also completely unsupervised, and thus prone to some systematic error, and therefore, it may not be the common or best practice to set hard filters based on these flags.

Additionally, chemicals in the present study were not selected based on their proven neurotoxicity according to the official *in vivo* approved methods. Therefore, it is unknown what is the accuracy of the bioassay to signal a neuroactive substance. Even if you can predict the activity in the *in vitro* assay, this does not necessarily reflect actual neurotoxicity signaling of the bioassay.

3.2.1 Sensitivity, specificity and accuracy

As explained in paragraph 3.1.3, toxicity classes were determined based on the distribution of experimental AC_{50} values within each *in vitro* assay focusing on neurodevelopment (See paragraph 3.1). Note that, since the toxicity

classes are based on the lower 25% and upper 75% of the AC_{50} distribution, there will always be three times more negatives than positives in each bioassay, which has implications for the PPV and NPV compared to the sensitivity and specificity of the models. The overall accuracy of the models was 89.8%, meaning that 89.8% of all toxicity classes (for all five endpoints, and for both the test and training dataset, including 'moderately' toxic chemicals under TRUE positive) were correctly predicted. Differences between the test and training dataset were very small, with an accuracy of 90.5% for the training dataset and 88.2% for the test dataset (Table 1). The overall sensitivity of all individual models (the percentage of correctly predicted "high toxicity" classes – true positives) for chemicals tested for the five subcategories, separately was 76.8% (training dataset: 77.3%, and test dataset: 75.6%), and the overall specificity of the models (the percentage of correctly predicted "low toxicity" classes – true negatives) 94.2% (training dataset: 94.9%, test dataset: 92.4%). This means that in general the models are better at preventing false negatives than false positives, which in turn implies that the models are conservative in turn leading to more precautious predictions. The average negative predictive value (NPV) for all models was 92.4% (training dataset: 92.6%, test dataset: 91.8%), implying that a high proportion of all cases with a negative QSAR prediction result truly exert low activity in the *in vitro* assays. The positive predictive value (PPV) for all models was 81.6% (training dataset: 83.5% ,test dataset: 77.1%), implying that 81.6% of all cases with a positive QSAR prediction result truly exert high activity in the *in vitro* assays.

Differences in specificity, sensitivity and accuracy occur when applying the models to the training dataset and test dataset (Figure 4), albeit very small (less than 3% between both dataset types). *In vitro* assay endpoints were clustered based on intended target subfamily. From all intended target subfamilies, the highest ***accuracy***, indicating overall model performance, was associated with the synaptogenesis model, when applied to the training dataset (94.9%), while the least accurate model was associated with neural network function: bursting activity, when applied to the test dataset (83.6%). The highest model ***sensitivity*** was calculated for the model based on the training dataset for *in vitro* assays related to synaptogenesis and neurite maturation (90.2%). However, the lowest model sensitivity was calculated for the model based on the test dataset of the same endpoint (69%), indicating that the model may be less powerful (with respect to preventing false negatives in the predictions, i.e., substances misclassified as not being active in *in vitro* assays related to neurodevelopment) when applying the model on data outside the training set. However, the model being less powerful in preventing false negatives is way better than the model being less powerful in preventing false positives, as in the prioritization and risk assessment of chemicals we prefer to have models that are too conservative over models that may not be conservative enough. ***Specificity*** (the models ability to predict true negatives) in general was high across all models (86%-98%), indicating that there are few false negatives, given that toxicity (AC_{50}s) was categorized in toxicity classes. A specificity of 100% indicates that there are no false negatives, indicating that – when applying this model as a first tier in risk assessment – no chemicals will be misclassified as non-toxic, meaning that no chemicals will be missed in follow-up research. The highest specificity (98%) was calculated based on models derived based on synaptogenesis in the test dataset (Figure 4). Additionally, for this specific endpoint (synaptogenesis), specificity for the model applied to the training set was also relatively high (96.6%). Although this model in particular may not be fully able to prevent false negatives, the model was able to better prevent false positives. Both the highest negative predictive value (NPV) and the highest positive predictive value (PPV) was calculated for synaptogenesis, when applying the model to the training dataset (96.6% and 90.2% for NPV and PPV, respectively). The NPV is strongly linked to the specificity of the model should ideally be 100% to rule out false negatives in first-tier risk assessment. Overall, NPVs for all developed QSARs were very high, but not 100%, implying that we cannot completely rule out false negatives by our QSARs. PPVs were high, but lower than the NPVs, implying that the models developed in this study were less likely to rule out false positives than false negatives, which may have implications for the number of times unnecessary follow-up research is suggested in higher tier risk assessment.

When applying the same model to *in vitro* assays within the test dataset (using chemicals that were not used as input in training the model), all model performance parameters (sensitivity, specificity, accuracy, NPV and PPV) were still

high (Table 1) compared to the other endpoints, implying that synaptogenesis is the neurotoxic endpoint most reliable to predict based on functional groups of compounds.

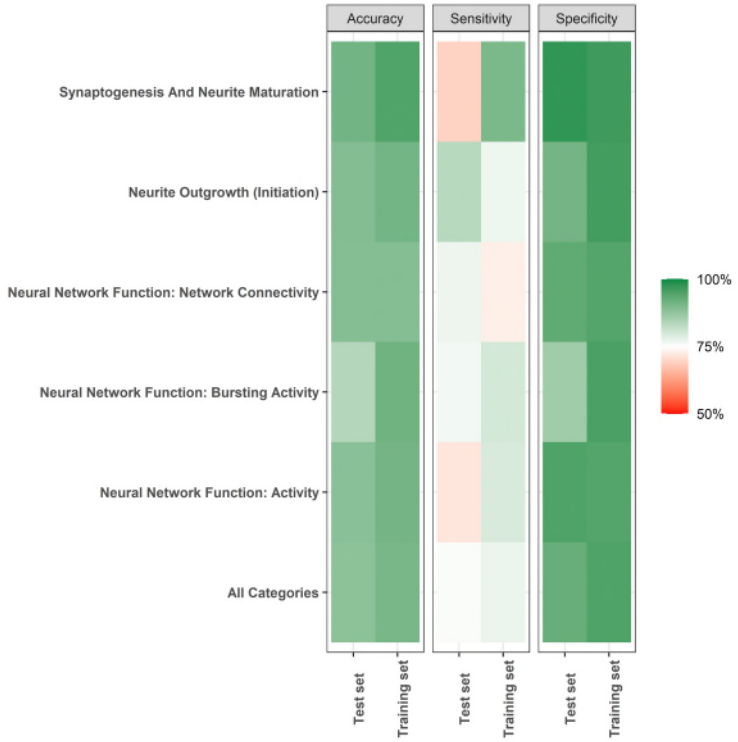

Figure 4. *Model performance indicators (Accuracy, sensitivity and specificity) based on the prediction of toxicity classes for the five separate endpoints (+ all categories, see paragraph 5.4), and when applying the model on the test and training datasets.*

Table 1: *model performance indicators (accuracy, sensitivity, specificity, NPV and PPV) per bioassay cluster and dataset type.*

	Accuracy		Specificity		Sensitivity		NPV		PPV	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
Neural network function: bursting activity	91,5%	83,7%	95,5%	86,0%	80,0%	76,5%	93,3%	92,0%	85,7%	63,4%
Neural network function: activity	90,9%	88,6%	94,5%	95,0%	79,6%	71,7%	93,6%	89,8%	82,1%	84,6%
Neural network function: network connectivity	89,1%	89,2%	94,4%	93,3%	72,9%	77,1%	91,3%	92,3%	81,2%	79,6%
Synaptogenesis and neurite maturation	94,9%	91,2%	96,6%	97,7%	90,2%	69,2%	96,6%	91,5%	90,2%	90,0%
Neurite outgrowth (initiation)	91,1%	89,3%	96,3%	90,9%	76,9%	83,3%	92,0%	95,2%	88,2%	71,4%
All categories	90,5%	88,2%	94,9%	92,5%	77,3%	75,6%	92,6%	91,8%	83,6%	77,2%

3.3 **Conclusions *in silico* neurotoxicity testing**

In the present study, we used multiple linear regression modelling to explore whether (developmental) neurotoxicity (AC_{50} values) can be predicted based on structural properties of chemicals, and to evaluate if sub-setting *in vitro* assay data based on intended target type subfamily aids in predicting neurotoxicity (bioactivity). In general, multiple linear regression explained a large part of the variance in toxicity when using structural fragments as explanatory variables, and also predicted AC_{50} values outside the training dataset very well. Although differences in model performance occur among subfamilies, all Q^2s when applying the models on data outside the training dataset for the five subfamilies were above 0, indicating a good model performance. When looking at the sensitivity, specificity and accuracy of the models when categorizing data in toxicity classes, we see that both the specificity and sensitivity were above 70%. In general, the specificity of the models was higher than the sensitivity, implying that the models performed better at preventing false positives than at preventing false negatives, implying that the models more often would misclassify a highly toxic compound as a low toxic compound, than the other way around. Nevertheless, sensitivity percentages of over 70% are still considered satisfactorily high. This means QSAR models can be used as a quick and cost-efficient alternative to determine bioassay activity in the five different bioassay types testing neuroactivity related effects. Additionally, we expect that there is a good chance the accuracy of detection of established human relevant and non-relevant neurotoxic substances by yet to be developed QSARS can be high. However, in the future, we recommend to test if these QSARs can predict true (systemic) neurotoxicity in whole organisms, rather than just predicting activity in *in vitro* assays associated with neurotoxicity.

4. *In vivo* models for neurotoxicity testing

4.1 Alternative *in vivo* models for (developmental) neurotoxicology

Due to the anticipated growth in global chemical production and the continuous introduction of new compounds into the market (Wilson & Schwarzman 2009), the conventional approach for testing neurotoxicity is insufficient. To address this issue, there is a need for new testing methods and alternative animal systems, such as non-mammalian models to meet the increasing demand of toxicity testing. The selection of the most promising alternative *in vivo* model for neurotoxicity testing relies on specific research objectives and the desired balance between simplicity and biological relevance. Therefore, it is imperative to evaluate the strengths and limitations of each model in order to make accurate decisions that align with the research objectives and ethical considerations. Table 2 presents a comparison of common *in vivo* models used for neurotoxicity testing along with their respective advantages and limitations.

Table 2. *Comparison of common in vivo models used for neurotoxicity testing*

Characteristics	**Model organism for neurotoxicity testing**				
	D. melanogaster *(fruit fly)*	***C. elegans*** *(nematode)*	***Planarians*** *(flatworms)*	***D. rerio*** *(zebrafish)*	***M. musculus*** *(mouse)*
Genetic similarity to humans	>60%	>50%	>50%	>70%	>90%
*Neuroregeneration**	Yes (limited)	Yes (limited)	Yes, (complete)	Yes (limited)	Yes (limited)
Screening throughput	low to medium	medium to high	medium to high	low to medium	very low
Generation time	10 – 14 days	3 – 5 days	3 - 7days	3 – 4 months	3 – 4 weeks
Adult size	3 mm	1 mm	10 – 30 mm	6 cm	10 cm
Growth medium	Solid	Solid/liquid	Liquid	Liquid	Cage
Ease to obtain individuals	High	High	Very high	Low	Very low
Maintenance	Low	Low	Low	High	Very high
Equipment cost	Low	Low	Low	High	Low
Advantages	Complex behaviour and genetic tools	Simple nervous system	Remarkable regenerative capabilities	Transparent embryos; high genetic and physiological similarity with humans	Close genetic and physiological similarity with humans; Well established behavioral assays
Limitations	Non mammalian model	Limited cognitive complexity	Limited cognitive complexity	Limited cognitive complexity	Ethical concerns; high maintenance cost
Key considerations	Genetic tractability	Cost effective and quick assays	Assessment of Neural regeneration	Rapid development	Regulatory acceptance
Ethical constraints	No	No	No	Not (upto 120 hpf)	Yes

D. melanogaster = Drosophila melanogaster ; C.elegans = Caenorhabditis elegans; D. rerio = Danio rerio; M. musculus = Mus musculus;
** Regrowth or repair of the nervous tissue by generating new neurons, axons, synapses, and glial cells.*

Hpf = hours post fertilisation

Sources: Mineta et al., 2003; Peterson et al., 2008; Giacomotto and Segalat 2010; Sandmann et al., 2011; Mirzoyan et al., 2019; Nagappan et al., 2020; Coupe and Bossing 2022; Tello et al., 2022.

Currently, a range of organisms are being used as alternative *in vivo* models including invertebrates such as *Drosophila melanogaster* or nematodes (*C. elegans*) and vertebrates (zebrafish). In our previous report, we presented different model systems that were then available for neurotoxicity testing, ranging from molecules (e.g., enzymes) and cells to intact model organisms such as nematodes (*C. elegans*) and zebrafish (Danio rerio). As such *D. melanogaster* and *C. elegans* are standard invertebrates models used in biology. It is worth noting that planarians were not part of the approaches described in our previous report (BTO 2020.035), signifying a deliberate departure to explore the unique advantages offered by these organisms. The selection of planarians for further investigation in this project is the result of a strategic decision based on a comprehensive assessment of this animal model. Furthermore, the departure from the methods presented in our previous report emphasises our ambition to broaden the scope of our investigations and to exploit the different strengths of the various model organisms. The rationale behind choosing planarians for in-depth investigation is their exceptional regenerative capabilities, which makes them valuable alternative animal models for neurotoxicity testing. Additionally, planarians can help gain insights into neurotoxic effects and neuroregeneration and offer simplicity, cost effectiveness, and relevance in assessing neuroactive substances (Hagstrom et al. 2016;). Their sensitivity to various chemicals falls within the sensitivity ranges or similar to those of the most commonly used invertebrate models (Wu and Li, 2018). More importantly, some of the genes are conserved between planarians and mammals and are not found either in *D. melanogaster* or *C. elegans* (Mineta et al., 2003). In addition, more toxicity endpoints can be analysed in planarians than other invertebrate *in vivo* models. Table 3. shows the toxicity endpoints that have been investigated using different invertebrate models. These findings support the validity of planarians for studying neurotoxicity and neurodevelopmental dysfunctions in response to exposure to neuroactive substances that may be present in aquatic system including water sources intended for human use. Specific initiatives, exist to probe the potential neurotoxicity of chemicals using alternative approaches (OECD, 2023), but there is still room for their improvement and for searching and validating alternative testing approaches.

Table 3. *Toxicity endpoints investigated using invertebrate models (Source: Wu et al., 2018).*

Toxicity endpoint	Cnidarians	Planarians	Annelids	Bivalves	Crustaceans
Teratogenicity	✓	✓	-	-	✓
Tumorigenicity	-	✓	-	✓	-
Genotoxicity	✓	✓	✓	✓	✓
Neurotoxicity	-	✓	✓	✓	✓
Behavioral toxicity	✓	✓	✓	✓	✓
Reproductive toxicity	✓	✓	✓	✓	✓
Biochemical disturbances	✓	✓	✓	✓	✓

4.2 Planarians as *in vivo* neurotoxicity bioassay

Planarians are free-living, non-parasitic metazoans belonging to the phylum Platyhelminthes (flatworms). The selection of planarians for our present study was based on their distinctive biological characteristics, practical advantages, exemption from ethical considerations and their potential to offer valuable insights into neurotoxicity. Hundreds of planarian species exist in marine, freshwater, and terrestrial environments worldwide (Ivankovic et al., 2019). Freshwater planarians are benthic communities and inhabit floating vegetation in unpolluted environments (Noreña et al., 2015; Wu and Li, 2018) (Figure.5). Their body is simple, ranging in size from 10-30 mm and possess a well-differentiated nervous system with neurons that share major neurotransmitters including dopamine, serotonin, octopamine, γ-aminobutyric acid (GABA), and acetylcholine and neuronal cell types with the vertebrate brain, allowing mechanistic insights into cellular effects and pathways that control specific behaviours (Lobo et al., 2012;

Ross et al., 2017; Zhang et al., 2019; Ireland et al., 2020). Their nervous system is dominated by a pair of ventral nerve cords that thicken in the head portion of the animal to form the brain or cerebral ganglion (Figure. 5). The head also contains the eyes and the auricles, which are the main sensory structures. Planarians possess protonephridia as excretory system and mesenchyme filling the space between the various organs of the body (Jordan and Verma, 2007). Planarians do not have specialized circulatory organs for oxygen transport. Oxygen is taken up through the body wall, while nutrients diffuse from the gut to the surrounding tissues.

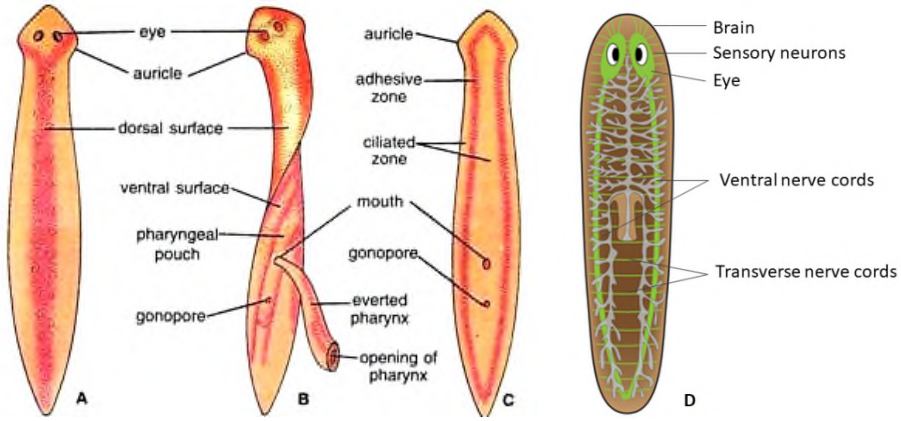

Figure 5: Planarian External features; *(A) dorsal surface, (B) body twisted to show a part of ventral surface, (C) ventral surface, (D) Nervous system* Source *(A-C):* https://ar.europeanwriterstour.com/images-2023/auricle-planaria

Planarians have received increasing scientific attention over the years due to their unique regenerating capability (Sánchez Alvarado, 2003; Hagstrom et al., 2016; Majid et al., 2022). The unique ability of planarians to regenerate their body, including the central nervous system (neurogenesis) within a few days, makes these animals an interesting candidate as a model for neurotoxicology and neurodegenerative studies. An important aspect of using planarians for neurotoxicity is that behavioral assays can be performed in parallel on both intact (adult) and amputated (regenerating/developing) animals to determine whether particular chemicals exert toxic effects on a developing or mature nervous system under varied controlled experimental conditions. This makes the planarian system an ideal complement to existing alternative animal models in toxicology, such as zebrafish and nematodes (Hagstrom et al. 2015, 2016). Planarians have been known to display distinct traits mediated by sensitive perception of environmental cues. For example, these exhibit a variety of abnormal body shapes, including signs of general sickness, constrictions, curling (C-shaped body), corkscrew-like, and displaying pharynx extrusion. Some body shapes have been associated with disturbances to specific neurotransmitter systems (Passarelli et al., 1999; Buttarelli et al., 2008), making body shape a potentially sensitive indicator of neurotoxicity. In addition, the planarian's CNS shares many of the same neurotransmitters and neurons with the mammalian brain including dopamine, serotonin, octopamine, GABA, and acetylcholine, providing researches the opportunity to gain potential insights into relevant mechanisms in humans (Cebrià et al. 2002; Mineta et al. 2003; Cebrià 2007; Hagstrom et al., 2016). For example, neuromuscular communication in planarians has been shown to be mediated by acetylcholine, as it is in humans (Nishimura et al. 2010; Hagstrom et al., 2016).

There is limited information about planarians being commonly used for testing drinking water relevant chemicals including the disinfection byproducts (DBPs). Although the planarians may not be a conventional choice for such assessments, the broad sensitivity of these organisms to a wide range of chemicals including environmental pollutants and pharmaceuticals, (Kapu & Schaeffer 1991; Rivera & Perich 1994; We et al., 2018; Hagstrom et la.,

2016; Majid et al., 2022) position them as potentially valuable tool for broader water quality assessments. Furthermore, the diversity among more than 400 species of freshwater planarians with range of sensitivities makes them highly valuable for water quality monitoring. The most common planarian species used in toxicological studies include *Dugesia Japonica, Schmidtea mediterranea,* and *Girardia tigrina*, (formerly *Dugesia tigrine*) (Ireland et al., 2020). *D. japonica* has been shown to be suitable for high-throughput screening (HTS) and *S. mediterranea* is the most popular planarian species for molecular studies as its fully annotated genome is available (Ireland et al., 2020; Morris et al., 2022). In addition, *S. mediterranea* is one of the best characterised organisms used in developmental biology, stem cell biology, ageing and neurotoxicological screenings (Newmark and Sánchez, 2002; Hagstrom et al., 2015, Hagstrom et al., 2016; Stevens et al., 2017, Fincher et al., 2018; Majid et al., 2021; Morris et al., 2022). Furthermore this model species is easy to handle, can be sufficiently multiplied in the laboratory and requires low maintenance cost. Therefore, in the present study, we propose the use of *S. mediterranea* as an *in vivo* model of neurotoxicity testing for water samples and drinking water relevant substances. For the purpose of this study, methods to evaluate neuroregeneration (see Appendix II) are considered out of scope in a water quality context and the focus will be on the methods used for neurotoxicity testing in planarians.

4.3 Methods for neurotoxicity testing using planarians

Here we provide an overview of the methodologies commonly employed for testing neurotoxicity in planarians. The assessment encompasses a range of metrics, including behavioral, morphological, cellular, and molecular parameters. Each of these methods allow the elucidation of the impact of the chemical substances on the neurobiology of planarians, offering a multifaceted approach to a neurotoxicity evaluation. Table 4 and 5 show a detailed breakdown of the methods used in these assessments.

Table 4. *Methods and organismal readouts to study neurophysiology and neurodevelopment in planarians.*

Test type	Endpoint	Method	Reference
Neurophysiology	Locomotion velocity	**Tracking of locomotion** Manual tracking of locomotion Automated tracking of locomotion using real-*time center of mass (COM) tracking system*	*Majid et al. 2022* *Hagstrom et al., 2015; 2016; Pag´an et al. 2012; Raffa et al. 2001*
Neurodevelopment	*Morphological development of head* *Anatomical development of CNS*	**Phenotypic screening** Evaluation of morphological aberrations in head, e.g., impaired head shape, eye defects, tissue damage. **Immunohistochemistry (Anti-SYNORF1)** - Analysis of re-establishment of cephalic ganglia - Quantification of brain sizes.	*Hagstrom et al.2016* *Majid et al., 2022* *Morris et al., 2022* *Majid et al., 2022*

Table 5. Pathways and markers to study neurotoxicity in planarians.

Pathway	Markers	Technique	References
Molecular and biochemical analysis	- Nervous system *Prohormone convertase 2 (pc2), Synaptogamin (syt), Glutamic acid decarboxylase (gad), Innexin-3 (inx3), Retinalhomeobox (rax), Orthopedia (otp)*	RT-PCR, in-situ hybridisation	*Hagstrom et al. 2015* *Balestrini et al. 2014* *Stevens et al. 2015*
	- Oxidative stress *Catalase (CAT), Superoxide Dismutase (SOD), Glutathione Peroxidase (GPX), Glutathione (GSH)*	RT-PCR	*Majid et al. 2022* *Van Roten et al. 2018* *Zhang et al. 2014, 2015* *Yuan et al. 2012* *Wu et al. 2012*
	Reactive oxygen species (ROS) *Glutathione content* *Superoxide dismutase activity*	Spectrophotometry and Confocal imaging	*Majid et al. 2022* *Bijnens et al. 2021* *Pirotte et al. 2015*
	- DNA damage *Xpa, rad51*	RT-PCR	*Jaenen et al. 2021* *Majid et al. 2022*
	- Programmed cell death (apoptosis) B-cell lymphoma 2 (*Bcl2), bax, caspase 3*	RT-PCR	*Majid et al. 2022*
	- Stem cell proliferation and maintenance *pcna, innexin -11, tumor protein p53, cdc23, cyclin-b*	RT-PCR	*Majid et al. 2022* *Stevens et al. 2015*
Cellular analysis	DNA damage Cell cycle responses (cell proliferation) Programmed cell death (Apoptosis)	Comet assay Anti-phospho-Histone H3 immunohistochemistry TUNEL assay	*Hwang et al. 2004* *Stevens et al. 2015* *Hagstrom et al. 2016* *Wouters et al. 2020* *Majid et al. 2022*

4.4 Expertise and equipment required for neurotoxicity assessment using planarians

Planarians are cost-effective and easy to culture in the laboratory settings. Table 6 presents a summary of the equipment for establishing, and maintaining the planarian colony and for conducting different neurological assessments. The cost involved in these techniques are not presented, as they may vary from actual prices (a detailed breakdown of the equipment's can be requested from the suppliers). Each of the different techniques outlined in Table 5 allows evaluation of toxicity in planarians. For example, evaluation of locomotor activity allows for assessing the dynamic behavioural patterns while as phenotypic screening, enables a broader examination of the potential morphological changes or abnormalities. Immunohistochemistry, spectrophotometry and RT-PCR can be used for cellular and molecular analysis. The integration of confocal imaging and fluorescence microscopy enhances the precision of observations by offering detailed, high-resolution images. Therefore, appropriate selection of these techniques, adapted to specific objective of evaluating water quality enables an effective approach in investigating neurotoxic effects of (drinking) water relevant contaminants. For a detailed procedure on housing, feeding, cleaning, culture expansion and the detailed description of each neurotoxicity testing method readers are referred to the publication of Merryman et al (2018) as it is out of the scope of this study.

Table 6. Expertise and equipment required to perform experiments on planarians.

Technique	Expertise	Equipment
Planarian culture	Experience in maintaining planarian culture.	Animal housing (plastic) containers, water, petri dishes, plastic transfer pipettes, paper towel or equivalent, temporary wastewater storage containers (2-4 L beakers) or additional housing containers, food (calf lever)
Locomotor velocity (LMV)	Knowledge about planarian behaviour.	CCD camera / COM tracker (excluding software cost)
Phenotypic screening	Knowledge about planarian morphology.	CCD camera, stereo microscope
Immunohistochemistry (IHC)	Knowledge of basic theory and scientific principles related to immunology, IHC and immunostaining; Laboratory expertise with running IHC; knowledge of quantitative image analysis techniques, software and data interpretation.	Fluorescence microscope NIS–Br software (Nikon Instruments)
Reverse transcription polymerase chain reaction (RT-PCR)	Knowledge of basic theory of cellular and molecular biology; expertise with running RT-PCR, including RNA/DNA isolation, software and data interpretation.	Instrument and software,
Spectrophotometry	Knowledge of basic theory and scientific principles related to spectrophotometry; expertise with analytical procedures and data interpretation.	Spectrophotometer
Confocal imaging	Knowledge of basic theory and scientific principles of confocal imaging; expertise with cell/tissue culture, molecular biology techniques.	Confocal microscope
Fluorescence microscope	Knowledge of basic theory of fluorescence microscopy; expertise with cell/tissue culture, molecular biology techniques.	Fluorescence microscope

4.5 Steps for neurotoxicity testing

The method for investigating the neurotoxic properties of water samples starts by subjecting the water samples containing a mixture of substances to chemical analysis to identify individual substances within the mixture. After identification of individual substances, the next step involves applying QSAR analysis to identify the potentially neurotoxic compounds within the mixture. This step helps in narrowing down and focusing on specific substances that may have neurotoxic properties. To validate the findings from QSAR analysis, specific neurotoxic substances identified by QSAR are subjected to testing on planarians. This step ensures practical validation of the neurotoxic effects indicated by QSAR analysis. If the QSAR analysis indicates neurotoxicity of the entire mixture, the confirmation can be adapted to include the original mixture rather than the individual substances. This adaptable confirmation process ensures a holistic understanding of the neurotoxic properties, and allows the selection of appropriate method for further analysis based on the results. In addition to monitoring neurophysiological and neurodevelopmental responses, the present methodology extends to a more comprehensive investigation, that includes mechanistic analysis and provides a multidimensional understanding of the potential neurotoxic effects of chemicals that may be present in water. In particular, the assessment of mechanistic aspects provides insights into the underlying biological processes that are influenced by neurotoxic substances.

4.6 Conclusion *in vivo* neurotoxicity testing

The use of planarians as a bioassay model for neurotoxicity testing offers several distinct advantages. Their unique ability to regenerate the central nervous system makes them particularly valuable for neutoxicological studies. The ease of culturing the planarians in the laboratory and their cost – effectiveness further enhance their suitability as animal models. Although planarians are not the conventional *in vivo* models for evaluating chemicals relevant to (drinking) water, their broad sensitivity to a variety of substances, including environmental pollutants and pharmaceuticals, emphasises their potential as a valuable tool for broader water quality assessment. The proposed *in vivo* approach for neurotoxicity testing of (drinking) water relevant chemicals if established could enhance our understanding of the effects of water quality on the nervous system. In addition, the adoption of this innovative method for neurotoxicity testing emerge as a proactive measure to preserve the integrity of drinking water and protect the health of vulnerable populations.

5. General Discussion

European Union legislation does not currently mandate the screening of neuroactive chemicals (Crofton et al., 2022). In terms of drinking water quality, the watch list that the drinking water companies are required to monitor (Directives (EU) 2020/2184) may be added or amended according to new scientific and legislative developments in order to ensure the highest standards and identify possible risks to drinking water at an early stage. The most recent update was made in 2020, when the European Drinking Water Directive was formally revised and new minimum standards for drinking water quality were introduced by the European Parliament (Directive (EU) 2020/2184, EU Commission, 2020). In the revised version, around thirty five additional substances including pharmaceutical products, microplastics, and endocrine disruptors were added to the list for further evaluation in terms of exposure and toxicity. To ensure that the drinking water companies are prepared for regulatory developments related to the inclusion of neuroactive substances, it is necessary to adopt alternative neurotoxicity testing methods that are rapid, cost-effective, and efficient in predicting the neurotoxicity of new chemicals in drinking water sources. This is of particular importance for vulnerable and susceptible populations groups, such as pregnant women, children and socially marginalised (Aschengrau et al., 2016; Dara and Drabovich, 2023). Ensuring the willingness of drinking water companies to adapt to regulatory changes and proactively assess the neurotoxicity of emerging substances not only aligns with the pursuit of high standards of water quality but also protects the health of individuals susceptible to potential effects.

QSARs have evolved into relatively robust and reliable quantitative methods for predicting the toxicity of chemicals and enable more accurate prediction of chemically induced neurotoxicity for a broader range of substances (Jiang et al., 2020). Due to its broad scope, this approach offers the potential for effective prioritisation and monitoring of contaminants for water quality purposes. The results presented in this report underline these statements, especially when predicting toxicity classes (high, medium and low neurotoxicity) rather than continuous toxicity endpoints. However, a standardised approach to QSAR is required to ensure that assessments of NT and DNT based on chemical structures (structural elements) are produced in a transparent, consistent and reliable manner. It is also important to recognise limitations, particularly in context of specificity, sensitivity and prediction of complex biological responses. Building on our previous study on QSAR and read-across modelling for water quality (Reus et al., 2022), this study introduces the innovative inclusion of the planarian model for the comprehensive assessment of neurotoxicity, which is expected to enable drinking water companies to enhance risk-based monitoring strategies. While QSAR model enhance the initial screening process and provide insights into potential neuroactive substances (Gadaleta et al., 2022), the outcome of the QSAR can be verified in the subsequent *in vivo* planarian bioassay. This serves as an important validation step for the holistic assessment of water quality. It takes into account the inherent

complexities of biological systems and ensures that the predictions made through QSAR align with actual neurotoxic effects.

According to the published literature, the planarian model appears to be suitable and cost-effective for monitoring water quality. Planarian research has provided toxicologists with useful new perspectives and adaptable methods for neurotoxicity assessment (Raffa et al., 2001; Hagstrom et al.2016; Jaenen et al. 2021; Majid et al., 2022). Furthermore, planarian models are useful in aquatic toxicology because of their sensitivity to a variety of pharmacological and toxicologic agents. Additionally, planarians share relevant biochemical and physiological susceptibilities to pollutants as mammals (Calevro et al., 1998). When standardised *in vivo* planarian models for neurotoxicity are fully established, the identification and differentiation of neuroactive substances in source or produced water can be further assured. This advancement holds the potential for multi-level implementation in water quality monitoring, ranging from basic morphological assessment to advanced molecular analyses. At the basic level, water quality monitoring could involve the use of a single planarian species such as *S. mediterranea* with a focus on morphological and behavioural endpoints. This approach, although simpler, provides a valuable initial screening tool for routine monitoring and allows assessment of gross (neuro) toxic effects in response to various water contaminants. At the intermediate level of implementation, the use of multiple planarian species will add another layer of diversity to the assessment. As different planarian species have different sensitivities to neurotoxic substances, providing a broader scope for monitoring, and capturing a range of responses that may be species-specific. For a more sophisticated and comprehensive analysis, advanced technologies such as transcriptome-sequencing (RNA-seq) can be integrated at the highest level of implementation to identify molecular targets for neurotoxicity. This level of analysis not only increases precision of neuroactive substances, but also enables a deeper mechanistic understanding of their effects.

In conclusion, in this study we demonstrate the applicability of *in silico* and an alternative *in vivo* model for the neurotoxicity testing to assess water quality. In addition to demonstrating applicability, detailed information on practical implementation in terms of cost, skills, robustness and throughput were presented. Both methods appear promising for neurotoxicological screening of water-relevant substances including those, for which data are insufficient as well as emerging chemicals and transformation products.

6. Future research

Building upon the findings of this report, several important recommendations for future research emerge, as follows.

1. *Pilot studies:* The report envisions pilot studies as a follow-up project in which the planarian model will be practically implemented at the laboratory and a test set comprising water-relevant substances and, in a later stage, water samples will be studied. This initiative aims to validate and refine the methods presented for neurotoxicity testing using planarian models.

2. *Collaborations with key organisations:* Neurotoxicity is a complex and broad field of research, which emphasizes the need for collaboration, both from a efficiency and scientific quality point of view. To establish the robustness and accuracy of the proposed innovative methods, collaboration with established organisations such as the NORMAN (Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances) provides a valuable opportunity to share the knowledge and integrate analytical methods for improved water quality assessment in a broader context. In addition, collaboration with experts from research institutes as initiated in the exploratory research with IRAS and RIVM should be continued.

3. *Address practical implementation*: Future research should not only focus on methodological advancements but also on the practical implementation of the developed techniques. This includes assessing the feasibility of integrating these methods into routine water quality monitoring practices and understanding their scalability for widespread application.

4. *Explore additional endpoints:* The inclusion of additional endpoints such as genotoxicity, carcinogenicity, and developmental toxicity in the pilot studies with the planarian model and the QSAR models represents an important opportunity for further investigation. This expanded scope allows for a more comprehensive understanding of the potential health impacts of water relevant contaminants, may lead to potentially predicting these endpoint through models, and provides insights to inform water quality managers about potential toxicity associated with contaminants, hence, initiate necessary actions such as improving treatment processes.

In summary, the proposed recommendations emphasise the importance of a multi-faceted approach in future research efforts by expanding test sets, exploring additional endpoints, fostering collaboration and addressing practical implications and challenges. This can significantly contribute to advancing the field of water quality assessment and developing proactive measures for sustainable water quality management.

References

- Agata, K., Umesono, Y. 2008. Brain regeneration from pluripotent stem cells in planarian. Philosophical Transactions of the Royal Society B: Biological Sciences 363(1500), 2071-2078.
- Apfeld J, Alper S., 2018. What Can We Learn About Human Disease from the Nematode C. elegans? Methods Mol Biol. 1706:53-75.
- Aschengrau, A., Janulewicz, P.A., 2016. White RF, Vieira VM, Gallagher LG, Getz KD, Webster TF, Ozonoff DM. Long-term Neurotoxic Effects of Early-life Exposure to Tetrachloroethylene-contaminated Drinking Water. Ann Glob Health. 2016 Jan-Feb;82(1):169-79.
- Baran, N., Rosenbom, A.E., Kozel, R. and Lapworth, D. 2022. Pesticides and their metabolites in European groundwater: Comparing regulations and approaches to monitoring in France, Denmark, England and Switzerland. Science of The Total Environment 842, 156696.
- Bellot, M., Bartolomé, H., Faria, M., Gómez-Canela, C. and Raldúa, D. 2021. Differential modulation of the central and peripheral monoaminergic neurochemicals by deprenyl in Zebrafish Larvae. Toxics 9(6), 116.
- Bencsik, A., Lestaevel, P., 2021. The Challenges of 21st Century Neurotoxicology: The Case of Neurotoxicology Applied to Nanomaterials. Front Toxicol. 18;3:629256.
- Bijnens K, Jaenen V, Wouters A, Leynen N, Pirotte N, Artois T, Smeets K., 2021. A Spatiotemporal Characterisation of Redox Molecules in Planarians, with a Focus on the Role of Glutathione during Regeneration. Biomolecules. 11(5):714.
- Bijnens, K., Jaenen, V., Wouters, A., Leynen, N., Pirotte, N., Artois, T. and Smeets, K., 2021. A spatiotemporal characterisation of redox molecules in planarians, with a focus on the role of glutathione during regeneration. Biomolecules 11(5), 714.
- Bilge, S., 2022. Neurotoxicity, types, clinical manifestations, diagnosis and treatment. Neurotoxicity-New Advances.
- Braun, G. and Escher, B.I., 2023. Prioritization of mixtures of neurotoxic chemicals for biomonitoring using high-throughput toxicokinetics and mixture toxicity modeling. Environment International 171, 107680.
- BTO 2020. Bioassays voor neuroactieve stoffen in water. BTO rapport; BTO 2020.035. KWR water research institute, The Netherlands.
- Busch, W., Schmidt, S., Kuhne et al., 2016. Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring Environ Toxicol Chem 2016;35:1887–1899. https://doi.org/10.1002/etc.3460
- C. Noreña, et al. 2015. Phylum Platyhelminthes. Ecology and General Biology: Thorp and Covich's Freshwater Invertebrates. Academic Press, Cambridge, Massachusetts, USA, pp. 181-203. https://doi.org/10.1016/B978-0-12-385026-3.00010-3.
- Caito S, Aschner, M., 2015. Neurotoxicity of metals. Handb Clin Neurol. 131:169-89.
- Cardenas-Iniguez C, Burnor E, Herting MM. 2022. Neurotoxicants, the Developing Brain, and Mental Health. Biol Psychiatry Glob Open Sci. 2(3):223-232.
- Cebria F, Kudome T, Nakazawa M, Mineta K, Ikeo K, Gojobori T, Agata K., 2002. The expression of neural-specific genes reveals the structural and molecular complexity of the planarian central nervous system. Mech Dev, 116:199-204.
- Cebrià, F., 2007. Regenerating the central nervous system: how easy for planarians! Development genes and evolution 217, 733-748.
- Cebrià, F., Nakazawa, M. , Mineta, K. , Ikeo, K. , Gojobori, T., Agata, K., 2002. Dissecting planarian central nervous system regeneration by the expression of neural-specific genes. Development, growth & differentiation, 44, pp.135–46

- Cebrià, F., Kudome, T., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T. and Agata, K. 2002., The expression of neural-specific genes reveals the structural and molecular complexity of the planarian central nervous system. Mechanisms of development 116(1-2), 199-204.
- Coupe D, Bossing T., 2022. Insights into nervous system repair from the fruit fly. Neuronal Signal. 13;6(1):NS20210051.
- Crofton, K. M., Bassan, A., Behl, M., Chushak, Y.G., Fritsche et al., 2022. Current status and future directions for a neurotoxicity hazard assessment framework that integrates *in silico* approaches. Comput Toxicol:100223.
- Crofton, K.M., Bassan, A., Behl, M., Chushak, Y.G., Fritsche, E., Gearhart, J.M., Marty, M.S., Mumtaz, M., Pavan, M. and Ruiz, P., 2022. Current status and future directions for a neurotoxicity hazard assessment framework that integrates in silico approaches. Computational Toxicology 22, 100223.
- Cunha-Oliveira, Teresa; Rego, Ana Cristina; Oliveira, Catarina R., 2008. Cellular and molecular mechanisms involved in the neurotoxicity of opioid and psychostimulant drugs. Brain Research Reviews. 58 (1): 192–208.
- Dara D, Drabovich AP. Assessment of risks, implications, and opportunities of waterborne neurotoxic pesticides. J Environ Sci (China). 2023 Mar;125:735-741.
- de Leeuw, V.C., Hessel, E.V., Pennings, J.L., Hodemaekers, H.M., Wackers, P.F., van Oostrom, C.T. and Piersma, A.H. 2020. Differential effects of fluoxetine and venlafaxine in the neural embryonic stem cell test (ESTn) revealed by a cell lineage map. Neurotoxicology 76, 1-9.
- Dingemans, M.M., Schütte, M.G., Wiersma, D.M., de Groot, A., van Kleef, R.G., Wijnolts, F.M. and Westerink, R.H., 2016. Chronic 14-day exposure to insecticides or methylmercury modulates neuronal activity in primary rat cortical cultures. Neurotoxicology 57, 194-202.
- Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W. and Kavlock, R.J. 2007. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicological sciences 95(1), 5-12.
- EFSA 2021. Development of Integrated Approaches to Testing and Assessment (IATA) case studies on developmental neurotoxicity (DNT) risk assessment. http://doi.org/10.2903/j.efsa.2021.6599
- EFSA 2022. European stakeholders' workshop on new approach methodologies (NAMs) for developmental neurotoxicity DNT) and their use in the regulatory risk assessment of chemicals. http://doi.org/10.2903/sp.efsa.2022.EN-7402
- Fitzgerald, J.A, Könemann, S., Krümpelmann, L., Županič, A., Vom Berg C., 2021 Approaches to Test the Neurotoxicity of Environmental Contaminants in the Zebrafish Model: From Behavior to Molecular Mechanisms. Environ Toxicol Chem. 2021 Apr;40(4):989-1006.
- Fritsche, E., Alm, H., Baumann et al., 2015. Literature review on in vitro and alternative Developmental Neurotoxicity (DNT) testing methods, EFSA Supporting Publication. 12(4):EN-778 186.
- Gadaleta, D., Spînu, N., Roncaglioni, A., Cronin, M.T. and Benfenati, E. 2022. Prediction of the neurotoxic potential of chemicals based on modelling of molecular initiating events upstream of the adverse outcome pathways of (developmental) neurotoxicity. International Journal of Molecular Sciences 23(6), 3053.
- Ghose, A.K., Viswanadhan, V.N. and Wendoloski, J.J., 1998. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. The Journal of Physical Chemistry A 102(21), 3762-3772.
- Giacomotto J, Ségalat L., 2010. High-throughput screening and small animal models, where are we? Br J Pharmacol. 160(2):204-16.
- Giordano, G. and Costa, L.G., 2012. Developmental neurotoxicity: some old and new issues. International Scholarly Research Notices. Article ID 814795
- Gouzoulis-Mayfrank, E., Daumann J., 2009. Neurotoxicity of drugs of abuse--the case of methylenedioxyamphetamines (MDMA, ecstasy), and amphetamines. Dialogues Clin Neurosci. 11(3):305-17.
- Grandjean, P. and Landrigan, P.J. 2006. Developmental neurotoxicity of industrial chemicals. The Lancet 368(9553), 2167-2178.

- Grummt, T., Braunbeck, T., Hollert, H. and Kramer, M. 2022. Hazard-based risk management of anthropogenic trace substances in drinking water to secure a long-term drinking water supply.
- Haas J, Manro J, Shannon H, et al., 2012 In Vivo Assay Guidelines. [Updated 2012 Oct 1]. In: Markossian S, Grossman A, Brimacombe K, et al., editors. Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK92013/
- Hagstrom, D., Cochet-Escartin, O. and Collins, E.M.S., 2016. Planarian brain regeneration as a model system for developmental neurotoxicology. Regeneration 3(2), 65-77.
- Hahn, M.E. and Sadler, K.C., 2020. Casting a wide net: use of diverse model organisms to advance toxicology, p. dmm043844, The Company of Biologists Ltd.
- Hair, J.F., Ringle, C.M. and Sarstedt, M., 2013. Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long range planning 46(1-2), 1-12.
- Hair, J.F., Ringle, C.M. and Sarstedt, M., 2013. Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long range planning 46(1-2), 1-12.
- Harry, J., Kulig, B., Lotti, M., Ray, D., Tilson, H. and Winneke, G., 2001. Neurotoxicity risk assessment for human health: principles and approaches. Environmental Health Criteria (223).
- Hernández-Jerez, A., Adriaanse, P., Aldrich, A., Berny, P., Coja, T., Duquesne, S., Focks, A., Marinovich, M., Millet, M., Pelkonen, O., Pieper, S., Tiktak, A., Topping, C., Widenfalk, A., Wilks, M., Wolterink, G., Crofton, K., Hougaard Bennekou, S., Paparella, M. and Tzoulaki, I., 2021. Development of Integrated Approaches to Testing and Assessment (IATA) case studies on developmental neurotoxicity (DNT) risk assessment. EFSA Journal 19(6), e06599.
- Hondebrink, L., Zwartsen, A. and Westerink, R.H., 2018. Effect fingerprinting of new psychoactive substances (NPS): What can we learn from in vitro data? Pharmacology & therapeutics 182, 193-224.
- Hubbs-Tait, L., Nation, J. R., Krebs, N. F., & Bellinger, D. C., 2005. Neurotoxicants, Micronutrients, and Social Environments: Individual and Combined Effects on Children's Development. Psychological Science in the Public Interest, 6(3), 57–121.
- Hubbs-Tait, L., Nation, J.R., Krebs, N.F. and Bellinger, D.C., 2005. Neurotoxicants, micronutrients, and social environments: Individual and combined effects on children's development. Psychological Science in the Public Interest 6(3), 57-121.
- Hwang, J.S,, Kobayashi, C., Agata K, Ikeo, K., Gojobori, T., Detection of apoptosis during planarian regeneration by the expression of apoptosis-related genes and TUNEL assay. Gene. 2004 May 26;333:15-25.
- Inoue, T., Kumamoto, H., Okamoto, K., Umesono, Y., Sakai, M., Alvarado, A.S. and Agata, K., 2004. Morphological and functional recovery of the planarian photo sensing system during head regeneration. Zoological science 21(3), 275-283.
- Jaenen, V., Fraguas, S., Bijnens, K., Heleven, M., Artois, T., Romero, R., Smeets, K. and Cebrià, F., 2021. Reactive oxygen species rescue regeneration after silencing the MAPK–ERK signaling pathway in Schmidtea mediterranea. Scientific reports 11(1), 881.
- Jenkins, J.E. and Roberts-Galbraith, R.H., 2023. Heterotrimeric G proteins regulate planarian regeneration and behavior. Genetics 223(4), iyad019.
- Jennifer E Jenkins, Rachel H Roberts-Galbraith., 2023. Heterotrimeric G proteins regulate planarian regeneration and behavior, Genetics, Volume 223, Issue 4.
- Jiang, C., Zhao, P., Li, W., Tang, Y. and Liu, G., 2020. *In silico* prediction of chemical neurotoxicity using machine learning. Toxicology research 9(3), 164-172.
- Kadry, H., Noorani, B., Cucullo, L., 2020. A blood–brain barrier overview on structure, function, impairment, and biomarkers of integrity. Fluids Barriers CNS 17, 69 (2020).
- Kamińska, K., Cudnoch-Jędrzejewska, A. A., 2023. Review on the Neurotoxic Effects of Doxorubicin. Neurotox Res 41, 383–397.

- Kapu, M.M., Schaeffer D,J., 1991. Planarians in toxicology. Responses of asexual Dugesia dorotocephala to selected metals. Bull Environ Contam Toxicol. 1991 47(2):302-7
- Klaassen, C.D., Amdur, M.O., 2013. Casarett and Doull's toxicology: the basic science of poisons, McGraw-Hill New York.
- Klassen, C.D., 1986. Distribution, excretion, and absorption of toxicants. Pp. 33-63 in Casarett and Doulls Toxicology. The Basic Science of Poisons. 3rd ed., CD Klassen, MO. Amdur, and J. Doull, eds. New York: Plenum Press.
- Kools, S., van Loon, A., Sjerps, R. and Rosenthal, L., 2019. De kwaliteit van bronnen van drinkwater in Nederland, KWR.
- Kropp, P.A., Bauer, R, Zafra, I., Graham, C., Golden, A., 2021. Caenorhabditis elegans for rare disease modeling and drug discovery: strategies and strengths. Dis Model Mech., 14 (8):
- Landrigan, P.J., Lambertini, L. and Birnbaum, L.S., 2012. A research strategy to discover the environmental causes of autism and neurodevelopmental disabilities. Environmental health perspectives 120(7), a258-a260.
- Legradi, J.B., Di, Paolo, C., Kraak, M.H.S. et al., 2018. An ecotoxicological view on neurotoxicity assessment. Environ Sci Eur 30, 46 2018.
- Lein, P., Silbergeld, E., Locke, P. and Goldberg, A.M., 2005. *In vitro* and other alternative approaches to developmental neurotoxicity testing (DNT). Environmental toxicology and pharmacology 19(3), 735-744.
- Lobo D, Beane WS, Levin M., 2012 Modeling planarian regeneration: a primer for reverse-engineering the worm. PLoS Comput Biol. 2012;8(4):e1002481. doi: 10.1371/journal.pcbi.1002481.
- Lopes, M.A., Meisel, A., Dirnagl, U., Carvalho, F.D., Bastos, Mde, L., 2008. Doxorubicin induces biphasic neurotoxicity to rat cortical neurons. Neurotoxicology. 2008 Mar;29(2):286-93.
- OECD/EFSA, 2016. OECD/EFSA Workshop on Developmental Neurotoxicity (DNT): the use of non-animal test methods for regulatory purposes. Available via: https://www.efsa.europa.eu/en/events/event/oecdefsa-workshop-developmental-neurotoxicity-dnt-use-non-animal-test
- Maciel, E.I., Jiang, C., Barghouth, P.G., Nobile, C.J., Oviedo, N.J., 2019. The planarian Schmidtea mediterranea is a new model to study host-pathogen interactions during fungal infections. Dev Comp Immunol. 93:18-27.
- Majid, S., Van Belleghem, F., Ploem, J.-P., Wouters, A., Blust, R. and Smeets, K., 2022. Interactive toxicity of copper and cadmium in regenerating and adult planarians. Chemosphere 297, 133819.
- Merryman, MS., Alvarado, AS., Jenkin, JC (2018). Editor: Jochen C.Rink, Planarian Regeneration: Methods and Protocols, Methods in Molecular Biology, Vol. 1774, http://doi.org/10.1007/978-1-4939-7802-1_5
- Mineta K, Nakazawa M, Cebria F, Ikeo K, Agata K, Gojobori T., 2003 Origin and evolutionary process of the CNS elucidated by comparative genomics analysis of planarian ESTs. Proc Natl Acad Sci U S A. 2003 Jun 24;100(13):7666-71.
- Mirzoyan. Z., Sollazzo, M., Allocca, M., Valenza, A.M., Grifoni, D., Bellosta, P., 2019. Drosophila melanogaster: A Model Organism to Study Cancer. Front Genet. 1;10:51.
- Moore, Keith; Dalley, Arthur., 2005. Clinically Oriented Anatomy (5th ed.). LWW. pp. 47. ISBN 0-7817-3639-0. A bundle of nerve fibers (axons) connecting neighboring or distant nuclei of the CNS is a tract.
- Morris, J., Bealer E.J, Souza, I.D.S., Repmann, L., Bonelli, H, Stanzione Iii, J.F, Staehle, M.M., 2022. Chemical Exposure-Induced Developmental Neurotoxicity in Head-Regenerating Schmidtea mediterranea. Toxicol Sci. 24;185(2):220-231.
- Nagappan, P.G., Chen, H., Wang, DY., 2020 Neuroregeneration and plasticity: a review of the physiological mechanisms for achieving functional recovery postinjury. Military Med Res 7, 30.
- National Research Council (US) Committee on Neurotoxicology and Models for Assessing Risk. Environmental Neurotoxicology. Washington (DC): National Academies Press (US); 1992. 1, Introduction: Defining the Problem of Neurotoxicity. Available from: https://www.ncbi.nlm.nih.gov/books/NBK234243/
- Newmark, P.A. and Alvarado, A.S., 2002. Not your father's planarian: a classic model enters the era of functional genomics. Nature Reviews Genetics 3(3), 210-219.

- Nishimura, K., Kitamura, Y., Taniguchi, T. and Agata, K., 2010. Analysis of motor function modulated by cholinergic neurons in planarian Dugesia japonica. Neuroscience 168(1), 18-30.
- Nguyen L, Rigo JM, Rocher V, et al., 2001. Neurotransmitters as early signals for central nervous system development. Cell and Tissue Research, 305(2):187–202.
- OECD, 2003. OECD guideline for the testing of chemicals. Test no 426. Developmental Neurotoxicity study. Available via: https://www.oecd.org/chemicalsafety/testing/15487898.pdf
- OECD, 2007. OECD guideline for the testing of chemicals. Test no 426. Developmental Neurotoxicity study. Available via: https://www.oecd-ilibrary.org/docserver/9789264067394-en.pdf?expires=1696264358&id=id&accname=guest&checksum=42557357E7523AD33691BCB626BC18D6
- OECD, 2023. Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery. Available via: https://one.oecd.org/document/ENV/CBC/MONO(2023)13/en/pdf
- Pagán, O.R., Baker, D., Deats, S., Montgomery, E., Tenaglia, M., Randolph, C., Kotturu, D., Tallarida, C., Bach, D. and Wilk, G., 2012. Planarians in pharmacology: parthenolide is a specific behavioral antagonist of cocaine in the planarian Girardia tigrina. International Journal of Developmental Biology 56(1-2-3), 193-196.
- Paparella, M., Bennekou, S.H. and Bal-Price, A., 2020. An analysis of the limitations and uncertainties of *in vivo* developmental neurotoxicity testing and assessment to identify the potential for alternative approaches. Reproductive Toxicology 96, 327-336.
- Pardridge, W.M., 1998. Recent advances in blood-brain barrier transport. Annu. Rev. Pharmacol. Toxicol. 28.25-39.
- Peterson, R.T., Nass, R., Boyd, W.A, Freedman, J.H, Dong, K, Narahashi T., 2008. Use of non-mammalian alternative models for neurotoxicological study. Neurotoxicology. 29(3):546-55
- Pistollato, F., de Gyves, E.M., Carpi, D., Bopp, S.K., Nunes, C., Worth, A. and Bal-Price, A. 2020. Assessment of developmental neurotoxicity induced by chemical mixtures using an adverse outcome pathway concept. Environmental Health 19(1), 23.
- Pronit, B., Whidul, H., J. J, Rajesh, K. K., Devasish, B., Rajesh, S. Y., 2022. Non-permitted food colorants induced neurotoxicity in cerebellum of rat brain, Drug and Chemical Toxicology, 45:6, 2852-2859
- Naidu, B. Biswas, I.R. Willett, J. Cribb, B. Kumar Singh, C. Paul Nathanail, F. Coulon, K.T. Semple, K.C. Jones, A. Barclay, R.J. Chemical pollution: A growing peril and potential catastrophic risk to humanityEnviron. Int., 156 (2021), Article 106616. https://doi.org/10.1016/j.envint.2022.107680
- Rabiasz A, Ziętkiewicz E., 2003. Schmidtea mediterranea as a Model Organism to Study the Molecular Background of Human Motile Ciliopathies. Int J Mol Sci. 2023 Feb 24;24(5):4472.
- Raffa, R.B., Holland, L.J. and Schulingkamp, R.J., 2001. Quantitative assessment of dopamine D2 antagonist activity using invertebrate (Planaria) locomotion as a functional endpoint. Journal of pharmacological and toxicological methods 45(3), 223-226.
- Reus, A., Pronk, T., ter Laak, T., Dingemans, M., 2020. Bioassays voor neuroactieve stoffen in water. H2O H2O-Online_200820_Bioassays_en_neurotoxiciteit.pdf (h2owaternetwerk.nl)
- Reus, A., Dingemans, M., Hoondert, R., Houtman, C., van der Oost, R., 2022. QSAR and read-across models for water quality. H2O Water Network. Available via: https://www.h2owaternetwerk.nl/vakartikelen/qsar-en-read-across-modellen-voor-waterkwaliteit
- Richardson, J.R., Fitsanakis, V., Westerink, RHS., Kanthasamy, A.G., 2019. Neurotoxicity of pesticides. Acta Neuropathol. (2019). 138(3):343-362. doi: 10.1007/s00401-019-02033-9.
- Rigdon, E., 2014. Rethinking Partial Least Squares Path Modeling: Breaking Chains and Forging Ahead. Long Range Planning 47.
- Robertson, S., 2022. What is Neurotoxicity?. News-Medical. Retrieved on July 31, 2023 from https://www.news-medical.net/health/What-is-Neurotoxicity.aspx.
- Rock, K.D. and Patisaul, H.B., 2018. Environmental mechanisms of neurodevelopmental toxicity. Current environmental health reports 5, 145-157.

- Rompolas P, Azimzadeh J, Marshall WF, King SM. , 2013. Analysis of ciliary assembly and function in planaria. Methods Enzymol;525:245-64.
- Ross, K.G, Currie K.W, Pearson B.J, Zayas R.M., 2017. Nervous system development and regeneration in freshwater planarians. Wiley Interdiscip. Rev. Dev. Biol 6, e266
- Ryan, N., Becker., R., 2017 A user's guide for accessing and interpreting ToxCast data, Bayer.
- Sandmann, T., Vogg, M.C., Owlarn, S., Boutros, M. and Bartscherer, K., 2011. The head-regeneration transcriptome of the planarian Schmidtea mediterranea. Genome biology 12, 1-19.
- Sarstedt, M., Ringle, C.M. and Hair, J.F., 2021. Handbook of market research, pp. 587-632, Springer.
- Schneider, J.S., 2023. Neurotoxicity and Outcomes from Developmental Lead Exposure: Persistent or Permanent? Environ Health Perspect. 131(8):85002
- Schriks, et al., 2012. Selection criteria to select *in vitro* bioassays for implementation and use. https://demeau-fp7.eu/system/files/results/D41%201_FINAL%2019102015%20(MS).pdf
- Schriks, M., Baken, K., Simon, E., Besselink, H., van der Linden, S., Kienle, C. and van der Burg, B., 2015. Selection criteria to select *in vitro* bioassays for implementation and use. Deliverable report D 41.
- Spencer, P.S., Lein, P., 2014. Neurotoxicity, Encyclopedia of Toxicology (Fourth Edition), Academic Press, 2024, Pages 727–740,
- Stevens, A.S., Pirotte, N., Plusquin, M., Willems, M., Neyens, T., Artois, T. and Smeets, K., 2015. Toxicity profiles and solvent–toxicant interference in the planarian Schmidtea mediterranea after dimethylsulfoxide (DMSO) exposure. Journal of Applied Toxicology 35(3), 319-326.
- Tello, J.A., Williams, H.E., Eppler, R.M., Steinhilb, M.L., Khanna, M., 2022. Animal Models of Neurodegenerative Disease: Recent Advances in Fly Highlight Innovative Approaches to Drug Discovery. Front Mol Neurosci. 15:883358.
- Tierney, K.B., 2011. Behavioural assessments of neurotoxic effects and neurodegeneration in zebrafish. Biochim Biophys Acta. 1812(3):381-9
- Turner, D.A. and Adamson, D.C., 2011. Neuronal-astrocyte metabolic interactions: understanding the transition into abnormal astrocytoma metabolism. Journal of Neuropathology & Experimental Neurology 70(3), 167-176.
- U.S. Environmental Protection Agency (EPA) 2014. US EPA TOXCAST DATA RELEASE ASSAY QUALITY SUMMARY OCTOBER 2014.
- U.S. EPA. Research and Guidance on Drinking Water Contaminant Mixtures. U.S. Environmental Protection Agency, Washington, D.C. accessed 10 October 2023 (available via: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=159132)
- Umesono, Y., Watanabe, K. and Agata, K., 1999. Distinct structural domains in the planarian brain defined by the expression of evolutionarily conserved homeobox genes. Development genes and evolution 209, 31-39.
- USEPA, 1998. Guidelines for Neurotoxicity Risk Assessment. Federal Register 63(93):26926-26954). https://www.epa.gov/sites/default/files/2014-11/documents/neuro_tox.pdf
- USEPA, 2014. US EPA TOXCAST DATA RELEASE ASSAY QUALITY SUMMARY OCTOBER 2014.
- using invertebrate (Planaria) locomotion as a functional
- van der Oost, R., Sileno, G., Suárez-Muñoz, M., Nguyen, M.T., Besselink, H. and Brouwer, A., 2017. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part I–model design and effect-based trigger values. Environmental toxicology and chemistry 36(9), 2385-2399.
- Van der Wolf, B. 2015. Seleciton criteria to select *in vitro* bioassays for implementation and use. (available via: https://demeau-fp7.eu/system/files/results/D41%201_FINAL%2019102015%20(MS).pdf
- Van Roten, A., Barakat, A.Z.A.-Z., Wouters, A., Tran, T.A., Mouton, S., Noben, J.-P., Gentile, L. and Smeets, K. 2018. A carcinogenic trigger to study the function of tumor suppressor genes in Schmidtea mediterranea. Disease models & mechanisms 11(9), dmm032573.

- Wang, R., Fu, Y. and Lai, L. 1997. A New Atom-Additive Method for Calculating Partition Coefficients. Journal of Chemical Information and Computer Sciences 37(3), 615-621.
- WHO/IPCS 2001. Neurotoxicity risk Assessment for Human Health: Principles and Approaches. Environmental Health Criteria 223. (available via: https://www.inchem.org/documents/ehc/ehc/ehc223.htm#_223110000)
- Wang, R., Fu, Y. and Lai, L. 1997. A New Atom-Additive Method for Calculating Partition Coefficients. Journal of Chemical Information and Computer Sciences 37(3), 615-621.
- Wilson, M.P. and Schwarzman, M.R. 2009. Toward a new US chemicals policy: rebuilding the foundation to advance new science, green chemistry, and environmental health. Environmental health perspectives 117(8), 1202-1209.
- Wlodkowic, D., Bownik, A., Leitner, C., Stengel, D., Braunbeck, T. 2022. Beyond the behavioural phenotype: Uncovering mechanistic foundations in aquatic eco-neurotoxicology. Science of The Total Environment. 829, 154584. https://doi.org/10.1016/j.scitotenv.2022.154584
- Wouters, A., Ploem, J.-P., Langie, S.A., Artois, T., Aboobaker, A. and Smeets, K. 2020. Regenerative responses following DNA damage–β-catenin mediates head regrowth in the planarian Schmidtea mediterranea. Journal of cell science 133(8), jcs237545.
- Wright, L., Pope, C. and Liu, J. 2009. Handbook of toxicology of chemical warfare agents, pp. 463-480, Elsevier.
- Wu L, Kalish BT, Finander B, Cao T, Jin G, Yahya T, Levy ES, Kukreja B, LaRovere ES, Chung JY, Lo EH, Brown-Whalen A, El Khoury J, Kaplan DL, Whalen MJ. 2022. Repetitive Mild Closed Head Injury in Adolescent Mice Is Associated with Impaired Proteostasis, Neuroinflammation, and Tauopathy. J Neurosci. 42(12):2418-2432.
- Wu, H.-Y., Shih, C.-L., Lee, T., Chen, T.-Y., Lin, L.-C., Lin, K.-Y., Chang, H.-C., Chuang, I.C., Liou, S.-Y. and Liao, P.-C., 2019. Development and validation of an analytical procedure for quantitation of surfactants in dishwashing detergents using ultra-performance liquid chromatography-mass spectrometry. Talanta 194, 778-785.

I Appendix

I.I Neurite outgrowth (initiation)

Neurite outgrowth is a fundamental process in the differentiation of neurons and a commonly used endpoint to study neuronal development and neuronal degeneration *in vitro*. Figure 6 shows the predicted effect concentrations (log_{10} AC_{50}) for eight bioassays focusing on neurite outgrowth[1,2,3,4,5,6,7,8]. plotted against the experimental effect concentrations from ToxCast, based on multiple linear regression analyses derived in the present study, taking the functional groups and topological parameters (See 3.1.1. *Data acquisition*) as explanatory variables. Two separate models were fitted: one where interaction between the explanatory variables were included (red dots, red text), and one where no interaction terms were included (blue dots, black text). In total, 94.55% of all individual predicted AC_{50}s lied within a factor of five of the observed AC_{50}s; 2.48% of the predicted datapoints were more than a factor five below the observed datapoints (underestimated), while 2.97% of all datapoints were more than a factor five above the observed data (overestimated). 0.50% of the predicted datapoints were a perfect fit, which may indicate overfitting of the model (Figure 6).

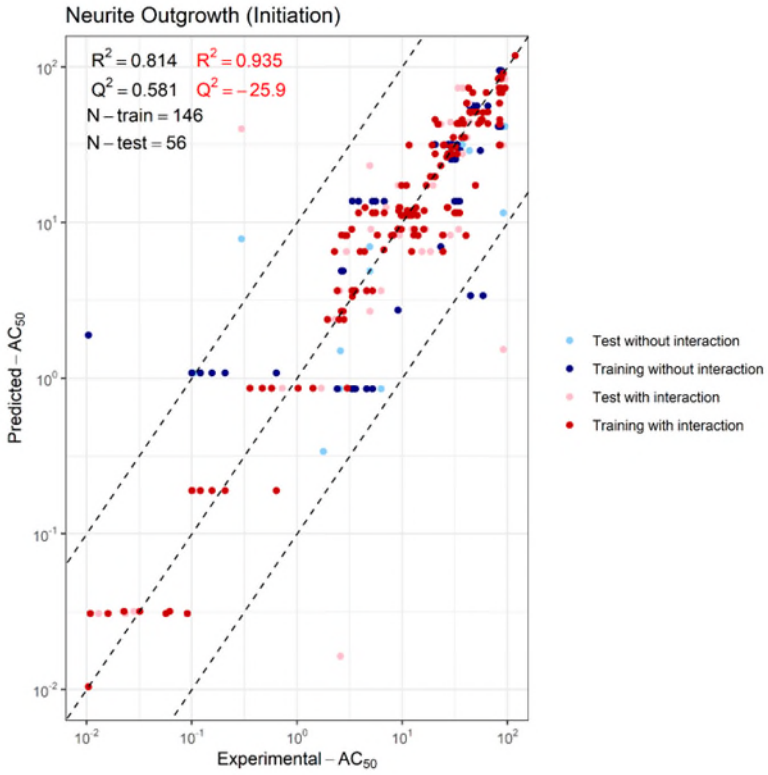

Figure 6. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus experimental toxicity values, including interaction terms (in red) and omitting interaction (in blue), for neurite outgrowth. Predictions were based on functional groups and topological parameters.*

In general, no major differences are apparent between the predictability of AC_{50} neurotoxicity values in the training dataset (on which the model is based) and the test dataset (containing substances on which the model has not been trained with) for the model without interaction terms. 96.91% of the predicted data points within the training dataset

[1] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_hN2_NOG_NeuronCount_loss
[2] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_hN2_NOG_BPCount_loss
[3] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_Cortical_NOG_NeuriteLength_loss
[4] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_hN2_NOG_NeuriteCount_loss
[5] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_Cortical_NOG_NeuriteCount_loss
[6] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_hN2_NOG_NeuriteLength_loss
[7] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_Cortical_NOG_BPCount_loss
[8] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCl_Cortical_NOG_NeuronCount_loss

were within a factor of five of the observed values, compared to 88.39% of the predicted data points in the test data set (Figure 6). However, the mean R^2 (0.814) of the model without interaction terms trained on the training dataset is similar to the mean Q^2 (0.235) when the models are applied to the test dataset. The Q^2, nevertheless, still indicates a good model performance for compounds outside the training dataset, given the rule of thumb that Q^2 values above zero indicate that the model has predictive relevance. Although the R^2 (based on values in the training dataset) for the model including interaction terms (0.935) is higher than for the model omitting interaction, its Q^2 (-25.9) indicates poor model performance/relevance for compounds outside the training dataset, which likely may due to the low sample size in the test dataset or high variability in the neurite outgrowth data per CAS number (indicated by the dashed lines in the figure).

I.II Synaptogenesis and neurite maturation

Synaptogenesis refers to the formation of synapses between neurons in the nervous system. The synaptogenesis and neurite maturation assay examines synapse formation. Figure 7 shows the predicted active effect concentrations (log_{10} AC_{50}) for eight bioassays focusing on synaptogenesis and neurite maturation[9,10,11,12,13,14,15,16], plotted against the experimental effect concentrations from ToxCast, based on multiple linear regression analysis, taking the functional groups and topological parameters (See 3.1.1. *Data acquisition*) as explanatory variables. Two separate models were fitted; one where interaction between the explanatory variables were included (in red), and one where no interaction terms were included (in blue). In total, 93.93% of all individual predicted AC_{50}s lied within a factor 5 of the observed AC_{50}s; 3.27% of the predicted datapoints were more than a factor five below the observed datapoints (underestimated), while 2.80 % of all datapoints were more than a factor five above the observed data (overestimated). None of the predicted datapoints were a perfect fit (Figure 7).

In general, no major differences are apparent between the predictability of AC_{50} neurotoxicity values in the training dataset (on which the model is based) and the test dataset (containing substances on which the model has not been trained) for the model with and without interaction terms. 92.9% of the predicted data points within the training dataset were within a factor of five of the observed values, compared to 96.49% of the predicted data points in the test data set (Figure 7). Although the percentage of predicted values within a factor five of the experimental values was lower in the training dataset than in the test dataset for both models, the mean R^2s (without interaction: 0.868, with interaction: 0.898) of the models for the training dataset are similar, but higher than the mean Q^2s (0.834 and 0.788, respectively) when the models are applied to the test dataset. This indicates that the average deviation of the predicted values from the experimental values is lower in the training dataset than in the test dataset.

[9] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_BPCount_loss
[10] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_NeuriteLength_loss
[11] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_NeuriteSpotCountPerNeuron_loss
[12] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_CellBodySpotCount_loss
[13] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_NeuriteCount_loss
[14] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_NeuronCount_loss
[15] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_SynapseCount_loss
[16] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Mundy_HCI_Cortical_Synap&Neur_Matur_NeuriteSpotCountPerNeuriteLength_loss

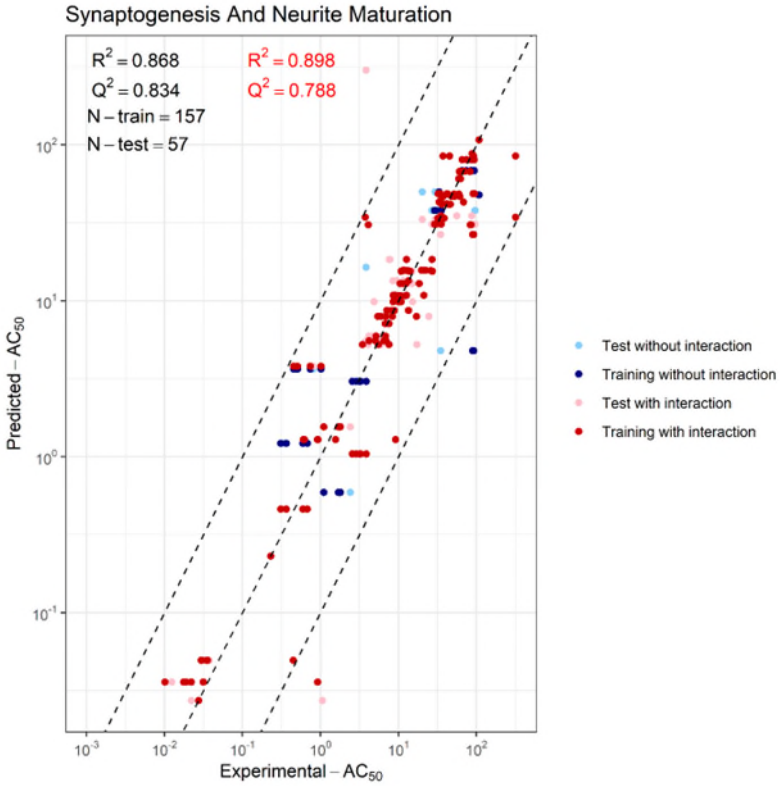

Figure 7. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus experimental toxicity values, including interaction terms (in red) and omitting interaction (in blue), for synaptogenesis and neurite maturation. Predictions were based on functional groups and topological parameters.*

I.III Neural network function: activity

Neuronal network activity (spikes of firing neurons) plays an important role in the development and maintenance of neuronal circuits. Figure 8 shows the predicted effect concentrations (Log_{10} AC_{50}s) for eight bioassays focusing neuronal activity[17,18,19,20,21,22,23,24], plotted against the experimental effect concentrations from ToxCast, based on multiple linear regression analysis, taking the functional groups and topological parameters (See 3.1.1. *Data acquisition*) as explanatory variables. Two separate models were fitted; one where interaction between the explanatory variables were included (in red), and one where no interaction terms were included (in blue). In total, 88.24% of all individual predicted AC_{50}s lied within a factor 5 of the observed AC_{50}s; 5.57% of the predicted datapoints were more than a factor five below the observed datapoints (underestimated), while 6.18% of all datapoints were more than a factor five above the observed data (overestimated). 0.09% of the predicted datapoints were a perfect fit, which may indicate overfitting of the model (Figure 8).

[17] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_bursting_electrodes_number_dn
[18] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_active_electrodes_number_dn
[19] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_burst_rate_dn
[20] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_firing_rate_mean_dn
[21] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_firing_rate_mean_up
[22] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_active_electrodes_number_up
[23] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_bursting_electrodes_number_up
[24] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_burst_rate_up

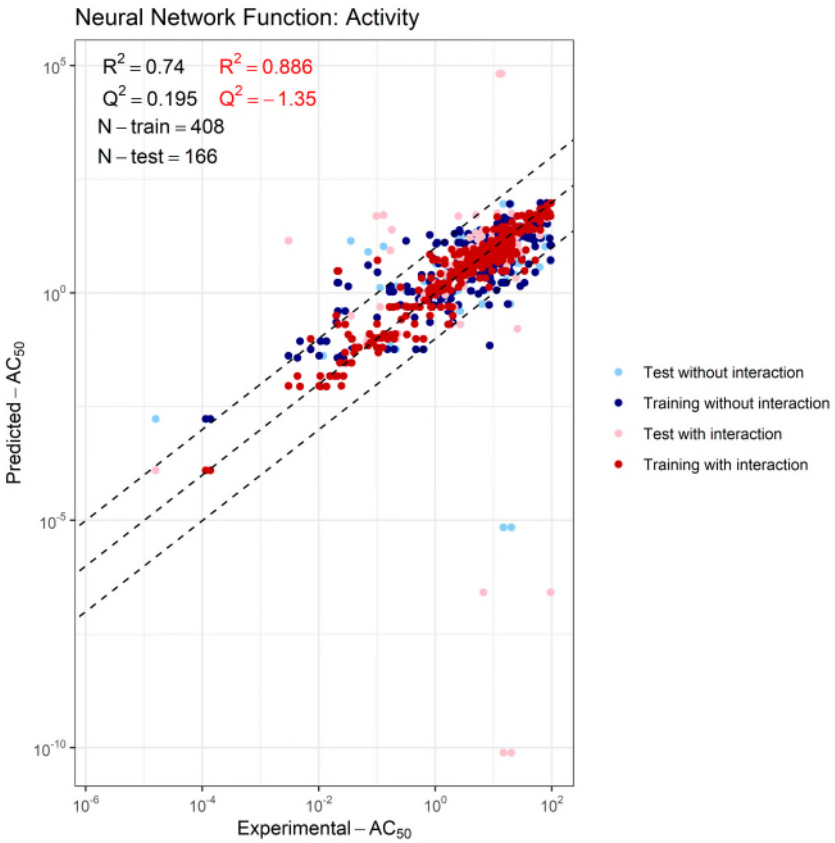

Figure 8. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus experimental toxicity values, including interaction terms (in red) and omitting interaction (in blue), for neural activity. Predictions were based on functional groups and topological parameters.*

In this case, major differences are apparent between the predictability of AC_{50} neurotoxicity values in the training dataset (on which the model is based) and the test dataset (containing substances on which the model has not been trained) for the model with and without interaction terms. 90.31% of the predicted data points within the training dataset were within a factor of five of the observed values, compared to 83.13% of the predicted data points in the test data set (Figure 8). Although the percentage of predicted values within a factor five of the test dataset for both models was relatively high, the mean Q^2s (0.0.195 and -1.35, for the model with and without interaction terms, respectively) were very low. The Q^2, nevertheless, indicates a sufficient model performance for compounds outside the training dataset in the model without interaction terms, as the Q^2 values above zero indicate that the model has predictive relevance. This requirement, however, was not met for the model including interaction terms, meaning that this model will not provide reliable predictions for compounds outside the training dataset.

I.IV Neural network function: bursting activity

Bursting activity in neurons refers to patterns of neural activity consisting of episodes of relatively fast spiking separated by intervals of quiescence. Figure 9 shows the predicted effect concentrations (Log_{10} AC_{50}s) for 18

bioassays focusing on bursting activity[25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42], plotted against the experimental effect concentrations from ToxCast, based on multiple linear regression analysis, taking the functional groups and topological parameters (See 3.1.1. *Data acquisition*) as explanatory variables. Two separate models were fitted; one where interaction between the explanatory variables were included (in red), and one where no interaction terms were included (in blue). In total, 90.18% of all individual predicted AC_{50}s lied within a factor 5 of the observed AC_{50}s; 4.55% of the predicted datapoints were more than a factor five below the observed datapoints (underestimated), while 4.65% of all datapoints were more than a factor five above the observed data (overestimated). 0.30% of the predicted datapoints were a perfect fit, which may indicate overfitting of the model (Figure 9).

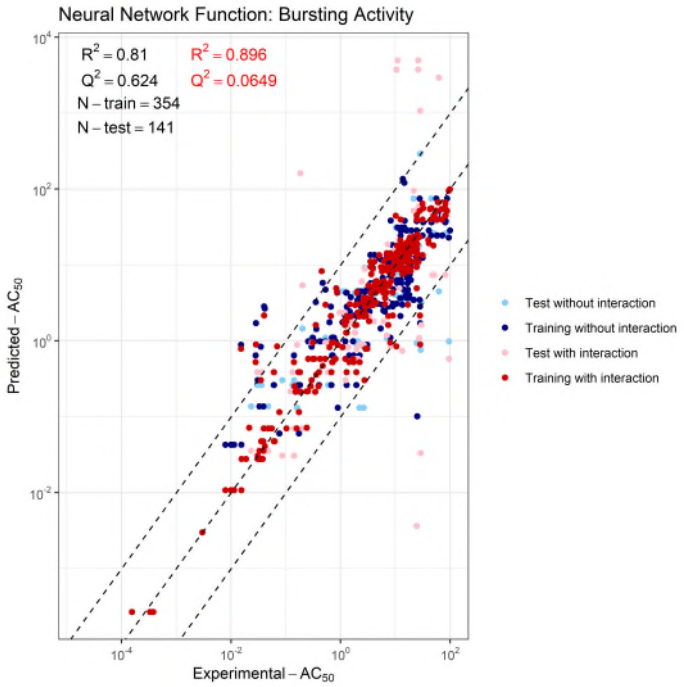

Figure 9. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus experimental toxicity values, including interaction terms (in red) and omitting interaction (in blue), for neural bursting activity. Predictions were based on functional groups and topological parameters.*

In general, no major differences are apparent between the predictability of AC_{50} neurotoxicity values in the training dataset (on which the model is based) and the test dataset (containing substances on which the model has not been trained) for the model with and without interaction terms. 93.64% of the predicted data points within the training

[25] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_spike_duration_mean_dn
[26] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_spike_percent_dn
[27] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_interspike_interval_mean_dn
[28] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_number_dn
[29] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_spike_number_mean_dn
[30] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_duration_std_dn
[31] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_correlation_coefficient_mean_dn
[32] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_peak_dn
[33] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_correlation_coefficient_mean_up
[34] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_spike_number_mean_up
[35] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_mutual_information_norm_dn
[36] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_spike_percent_up
[37] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_per_network_spike_interspike_interval_mean_up
[38] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_mutual_information_norm_up
[39] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_peak_up
[40] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_duration_std_up
[41] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_network_spike_number_up
[42] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_spike_duration_mean_up

dataset were within a factor of five of the observed values, compared to 83.68% of the predicted data points in the test data set (Figure 9). However, although the percentage of predicted values within a factor five of the test dataset for both models was relatively high, the mean Q^2 for the model with interaction terms (0.065) was very low. The Q^2s of both models, nevertheless, still indicate a sufficient model performance for compounds outside the training dataset, as both the Q^2 values above zero indicate that the model has predictive relevance.

I.V Neural network function: network connectivity

Figure 8 shows the predicted effect concentrations (Log_{10} AC_{50}s) for eight bioassays focusing on neural network connectivity[43,44,45,46,47,48,49,50], plotted against the experimental effect concentrations from ToxCast, based on multiple linear regression analysis, taking the functional groups and topological parameters (See 3.1. *Data acquisition*) as explanatory variables. Two separate models were fitted; one where interaction between the explanatory variables were included (in red), and one where no interaction terms were included (in blue). In total, 88.55% of all individual predicted AC_{50}s lied within a factor 5 of the observed AC_{50}s; 4.79% of the predicted datapoints were more than a factor five below the observed datapoints (underestimated), while 6.66% of all datapoints were more than a factor five above the observed data (overestimated). None of the predicted datapoints were a perfect fit (Figure 10).

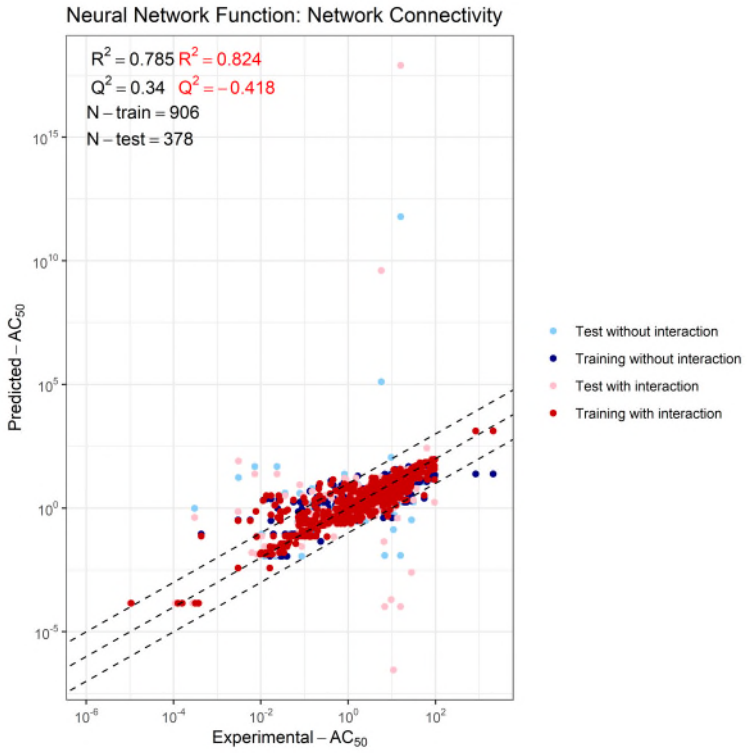

Figure 10. *Predicted toxicity (AC_{50} in µM) by the multiple linear regression model versus experimental toxicity values, including interaction terms (in red) and omitting interaction (in blue), for network connectivity. Predictions were based on functional groups and topological parameters.*

[43] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_bursting_electrodes_number_dn
[44] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_active_electrodes_number_dn
[45] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_burst_rate_dn
[46] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_firing_rate_mean_dn
[47] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_firing_rate_mean_up
[48] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_active_electrodes_number_up
[49] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_bursting_electrodes_number_up
[50] https://comptox.epa.gov/dashboard/assay-endpoints/CCTE_Shafer_MEA_dev_burst_rate_up

Although the percentage of predicted values within a factor five of the test dataset for both models was relatively high, the mean Q^2s (0.0.195 and -0.418, for the model with and without interaction terms, respectively) were quite low. The Q^2, nevertheless, indicates a sufficient model performance for compounds outside the training dataset in the model without interaction terms, as the Q^2 values above zero indicate that the model has predictive relevance. This requirement, however, was not met for the model including interaction terms, which is due to a limited subset of big predicted variations for chemicals with large experimental AC_{50}s (higher than 1 µM).

II Appendix

Planarian Brain Regeneration

Planarians, including *S. mediterranea*, regain full morphology after decapitation and the formation of the a new head (cephalic ganglia) occurs within a week (Cebria et al., 2002; Sandmann et al., 2011) (Figure 11). The high regenerative capacity is derived from a large pool of adult somatic stem cells, called neoblasts. Neoblasts known for their pluripotency and plasticity, account for 20-30% of all the animal's cells and are distributed throughout the mesenchyme (except in front of the photoreceptors and pharynx) (Morgan, 1988; Reddien and Sánchez Alvarado, 2004; Baguñà and Romero, 1981, Zhu and Pearson, 2016, Ivankovic, 2019). Neoblasts are the only mitotically active cells and generate the 30-40 different cell types found in these organisms. In intact planarians these stem cells replace cells lost to normal physiological turnover, whereas, in amputated animals, they give rise to the regeneration of blastema and enable them to grow lost tissue, and maintain tissue integrity indefinitely (Davies et al., 2017). This allows direct comparison of adult and regenerating animals without possible complications due to variability of genetic factors (Hagstrom et al., 2015). When an animal is amputated, the trunk piece can regenerate a brain, including visual neurons. The brain function gets fully restored in one week after amputation (Inoue et al., 2002; Morris et al., 2022). The process of neuroregeneration can be divided into five steps based on the histological observations and gene expression patterns (Agata and Umesono, 2008):

(1) formation of anterior blastema after wound closure
(2) formation of brain rudiment in the anterior blastema

The first two steps occur within 24 hours after amputation and are difficult to distinguish,

(3) pattern formation
(4) formation of neural network, and
(5) the recovery of function

A detailed description of the mechanism of brain formation is out of the scope of this study.

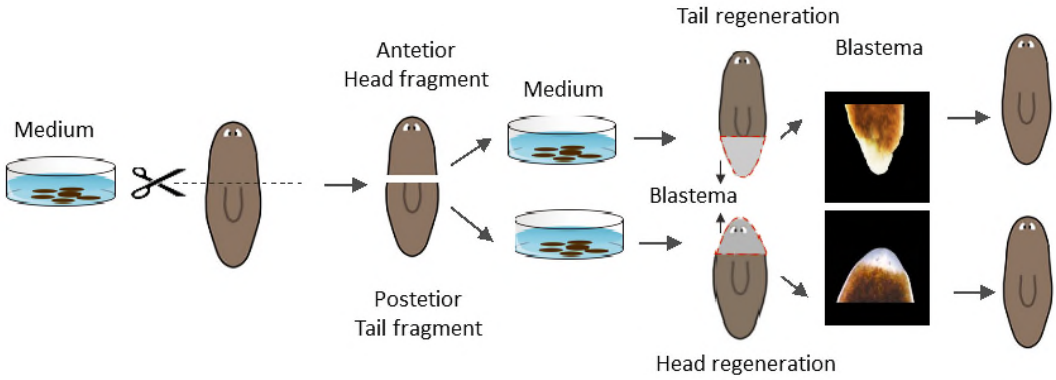

Figure 11. Regeneration in planarian Schmidtea mediterranea: *Regeneration involves the formation of new tissue at the wound site via proliferation of dividing neoblasts (a homogeneous stem cell population), resulting in the formation of blastema (white unpigmented part; indicated by red outline) and the remodelling of pre-existing tissues to restore symmetry and proportion (morphallaxis). It has been shown that planarians regeneration is the result of carefully orchestrated communication both within the blastema and between the newly regenerating tissues and the old existing tissues. Image source: Majid et al. (2022).*