INTEGRATING *IN SILICO* AND *IN VITRO* APPROACHES: CYTOTOXICITY AND ENZYMATIC ACTIVITY OF XENOBIOTICS IN DIFFERENT FISH CELL LINES

by
Elif Nagihan Kahraman
B.Ed. in Science Education, Boğaziçi University, 2015

Submitted to the Institute of Environmental Sciences in partial fulfillment of
the requirements for the degree of
Master of Science
in
Environmental Sciences

Boğaziçi University
2018

INTEGRATING *IN SILICO* AND *IN VITRO* APPROACHES: CYTOTOXICITY AND ENZYMATIC ACTIVITY OF XENOBIOTICS IN DIFFERENT FISH CELL LINES

APPROVED BY:

Prof. Dr. Melek TÜRKER SAÇAN . . . . . .
Thesis Advisor

Prof. Dr. Safiye SAĞ ERDEM . . . . . .

Prof. Dr. Nilsun İNCE . . . . . .

DATE OF APPROVAL: 10/09/2018

ACKNOWLEDGMENTS

Firstly and most importantly I would like to appreciate the value of ongoing effort and motivation my thesis advisor Prof. Dr. Melek Türker Saçan provided me through my study. She never stopped believing me even when I did not believe in myself. Also I want to thank Prof. Saçan for letting me to study at the Ecotoxicology and Chemometrics Laboratory where I have learned so much and achieved beyond what I expected with her guidance. I would be happy to present my sincere thanks to the jury members Prof. Dr. Nilsun İnce and Prof. Dr. Safiye Sağ Erdem for their helpful suggestions and highly motivating comments.

A very special thanks to Assoc. Prof. Ebru Kaya and Assist. Prof. Ebru Zeynep Muğaloğlu for all the things that they taught me and for their reference letters that helped me for my acceptance to the Institute of Environmental Sciences.

I would like to thank Ayşe Defne Şahin, Dr. Gülçin Tuğcu and Selen Gökçe for all enjoyable and informative working hours at laboratory. It was a pleasure working with you. Also my friend Ezelnur Aydemir deserves my deepest gratitude for getting my mind out of all the difficulties I faced during my studies. A late but fully appreciated motivation during my thesis study came from my friends at Marmara University so I wish to thank all of them, especially to Gizem Ayermolalı.

Lastly, I want to express my greatest thanks to the most important people in my life, to my family. My beloved siblings Barboros Hamdi and Hatice Nuriye bring my life constant happiness and provide all the support I needed not only for my thesis but for everything. I want to use my chance here to state again that I love you so much guys. My mother, Hanife Kahraman, I can never thank enough to compensate her dedication and her support through all issues in my life. She is and will always be the one who could stand out against the hardest and tough times with continuous love and patience. Also I want to thank all the other family members, especially my aunt Hatice Kahraman, who did not refrain to support me during my thesis. Finally, I wish my words to reach my father, Yusuf İzzettin Kahraman, bringing him some peace and honour regarding his daughter.

I gratefully appreciate the financial supports of the Scientific and Technological Research Council of Turkey (TUBITAK) (Project No: 214Z225).

ABSTRACT

INTEGRATING *IN SILICO* AND *IN VITRO* APPROACHES: CYTOTOXICITY AND ENZYMATIC ACTIVITY OF XENOBIOTICS IN DIFFERENT FISH CELL LINES

In the present study, five Quantitative Structure-Toxicity Relationship (QSTR) models were generated. Three QSTR models were developed for the cytotoxcity (pEC_{50}) of diverse chemicals to *Poeciliopsis lucida* hepatocarcinoma cell line (PLHC-1) measured with three assays, namely ethoxyresorufin-O-deethylase enzyme (EROD), neutral red (NR) and 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assays. EROD, NR and MTT assays reflects enzymatic activity, lysosomal and mitochondrial activity, respectively. The other two QSTR models were generated using the NR data for goldfish (GFS), *Carassius auratus*, scale tissue and fathead minnow (FHM), *Pimephales promelas*, epithelial cell line. Descriptors appearing in each model were selected via the tools implemented in QSARINS 2.2.1 software. All QSTR models were generated in line with the Organization of Economic Co-operation Development (OECD) principles and validated both internally and externally. $pEC_{50,EROD[PLHC-1]}$ model had one descriptor, $pEC_{50,NR[GFS]}$ had two and each of $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[FHM]}$ and $pEC_{50,MTT[PLHC-1]}$ models had three descriptors from DRAGON. The external predictivity of the generated QSTR models were tested using structurally diverse chemicals with no experimental cytotoxicity data. Structural coverage of the $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$, $pEC_{50,MTT[PLHC-1]}$, $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[FHM]}$ models for the external set compounds were 98.6%, 95.0%, 98.3%, 92.1% and 89.1%, respectively. A moderate/strong correlation was observed between the experimental *in vivo* and predicted *in vitro* values for the external set chemicals, except fathead minnow. The generated QSTR models may provide an initial, rapid screening and prioritization of these diverse chemicals for the acute fish toxicity assessment and reduce the need for extensive *in vivo* toxicity testing.

ÖZET

IN SILICO VE *IN VITRO* YAKLAŞIMLARIN BÜTÜNLEŞTİRİLMESİ: KSENOBİYOTİKLERİN FARKLI BALIK HÜCRE DİZİLERİ ÜZERİNDEKİ SİTOTOKSİSİTESİ VE ENZİMATİK AKTİVİTEYE ETKİSİ

Bu çalışmada, beş kantitatif yapı-toksisite ilişkisi (QSTR) modeli geliştirilmiştir. Çeşitli kimyasalların *Poeciliopsis lucida* hepatokarsinom hücre dizisi (PLHC-1) kullanılarak üç farklı test - etoksiresorufin-O-deetilaz enzimi (EROD), neutral red (NR) ve 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolyum bromit (MTT) ile sitotoksisitesi ölçülmüş ve bu veriler kullanılarak üç QSTR modeli geliştirilmiştir. EROD, NR and MTT testleri sırasıyla enzimatik aktivite, lizozomal hasar ve mitokondriyal aktiviteyi göstermektedir. Diğer iki QSTR modeli ise akvaryum balığı (GFS), *Carassius auratus*'un pul dokusuna ve fathead minnow (FHM), *Pimephales promelas*'ın epitel hücre dizisine ait NR verisi kullanılarak geliştirilmiştir. Modellerdeki tanımlayıcılar QSARINS 2.2.1 yazılımında bulunan araçlarla seçilmiştir. Tüm QSTR modelleri Ekonomik İşbirliği ve Kalkınma Örgütü'nün (OECD) belirlediği ilkelerle uyumlu olacak şekilde geliştirilmiş ve dâhili ve harici olarak doğrulanmıştır. $pEC_{50,EROD[PLHC-1]}$ modeli bir tanımlayıcı, $pEC_{50,NR[GFS]}$ modeli iki ve $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[FHM]}$ ve $pEC_{50,MTT[PLHC-1]}$ modellerinin her biri üçer tanımlayıcılıdır ve bu tanımlayıcılar DRAGON 6.0 yazılımı ile hesaplanan tanımlayıcılardır. Geliştirilen QSTR modellerinin tahmin performansı, farklı yapıda ve deneysel sitotoksisite verisi bulunmayan kimyasalların oluşturduğu birer harici set ile test edilmiştir. $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$, $pEC_{50,MTT[PLHC-1]}$, $pEC_{50,EROD[PLHC-1]}$ ve $pEC_{50,NR[FHM]}$ modellerinin harici set kimyasallarını yapısal olarak kapsama yüzdeleri sırasıyla %98.6, %95.0, %98.3, %92.1 ve %89.1 olarak bulunmuştur. Fathead minnow modeli dışında, harici set kimyasalları için deneysel *in vivo* ve tahmin edilen *in vitro* değerleri arasında orta veya güçlü korelasyon gözlemlenmiştir. Bu çalışmada geliştirilen QSTR modelleri, çeşitli kimyasalların akut balık toksisitelerinin değerlendirilmesinde bir hızlı öntarama ve önceliklendirme sağlayabilir ve kapsamlı toksisite testlerine olan ihtiyacı azaltabilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation	Unit
$^{1}X^{v}$	Molecular Connectivity Index	
3D-MoRSE	Molecule Representation of Structures Based on Electron Diffraction	
CYP1A	Cytochrome P4501A	
E	Energy	eV
E_{aq}	Energy in Aqueous Phase	eV
EC_{50}	Concentration of a Chemical that Causes %50 Effect on a Target Organism	mM
$EC_{50,EROD}$	Concentration of Toxicant that causes 50% Increase in Concentration of Resorufin	mM
$EC_{50,MTT}$	Concentration of Toxicant that Causes 50% Reduction in the Cleavage of MTT Dye	mM
$EC_{50,NR}$	Concentration of toxicant that Causes 50% Reduction in the Uptake of NR Dye	mM
E_{HOMO}	Energy of the Highest Occupied Molecular Orbital	eV
E_{LUMO}	Energy of the Lowest Unoccupied Molecular Orbital	eV
G	Geometric Distance Matrix	
F	Fischer Statistics	
X5v	Valence Connectivity Index of Order 5	
GATS2v	Geary Autocorrelation of Lag 2 Weighted by van der Waals Volume	
*h**	Critical Hat Value	
^{k}D	Molecular Profile Descriptors	
LC_{50}	Concentration of a Chemical that Causes %50 Lethality a Target Organism	mM
log K_{ow}	*n*-Octanol-Water Partition Coefficient	
Mor28e	Signal 28 / Eeighted by Sanderson Electronegativity	
MTT	3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide	
NaasC	Number of Atoms of Type aasC Atom-type	

PLHC-1	Topminnow (*Poeciliopsis lucida*) Hepatoma Cell Line
pT	Negative Logarithm of Toxic Concentration
R3m+	R Maximal Autocorrelation of Lag 3 / Weighted by Mass
TCDD	2,3,7,8-Tetrachlorodibenzodioxin

Abbreviation	**Explanation**
AD	Applicability Domain
AhR	Aryl-hydrocarbon Receptor
APIs	Active Pharmaceutical Ingredients
CATS	Chemically Advanced Template Search
CCC	Concordance Correlation Coefficient
DDT	Dichloro Diphenyl Trichloroethane
DLS_04	Modified Drug-Like Score from Chen et al. (7 rules)
DP07	Molecular Profile No. 7
ECOSAR	ECOlogical Structure Activity Relationship
ECHA	European Chemicals Agency
EROD	Ethoxyresorufin-O-Deethylase
FHM	Fathead Minnow
GETAWAY	Geometry, Topology and Atom-Weight Assembly
GFS	Goldfish Scale Tissue
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MCDM	Multiple Criteria Decision Making
MIM	Molecular Influence Matrix
MLR	Multiple Linear Regression
NPAHs	Nitrated Polycyclic Aromatic Hydrocarbons
NR	Neutral Red
OECD	Organization of Economic Co-operation Development
OLS	Ordinary Least Squares
PAHs	Polycyclic Aromatic Hydrocarbons
PCBs	Polychlorinated Biphenyls
PCDDs	Polychlorinated Dibenzodioxins
PPCPs	Pharmaceuticals and Personal Care Products
QSA/TR	Quantitative Structure-Activity/Toxicity Relationship
US EPA	The United States Environmental Protection Agency

1. INTRODUCTION

The release of chemicals into the environment is of great concern regarding their effects on living organisms. However, it is not practical to use *in vivo* testing to characterize the toxicological effects of millions of chemicals in use and/or produced (Armitage et al., 2014). This is, indeed, not only related to the high number of chemicals or lack of useful methods used for the determination of toxicity but also there are also some other considerations related to this difficulty. One of the most critical problems with *in vivo* procedures is the use of large numbers of test animals.

For environmental toxicology analyses, fish is one of the most widely used groups among vertebrates. It is a large group containing approximately 20,000 species expanding over a broad range of aquatic niches (Bols et al., 2005). Therefore, its usage in toxicology studies can provide a comprehensive assessment of toxicants' effect on the aquatic environment. Furthermore, the effects of toxicants to fish can give a general understanding of their effects on human. However, the use of whole animal for cytotoxicity assays, i.e. *in vivo* analysis, results in unethical treatment of a high number of animals. For example, OECD 203 (Organization of Economic Co-operation Development, OECD, 1992) assay prescribes 42-60 fish per test for acute toxicity. Therefore, it is attempted to decrease the number of animal tests for the determination of environmental risks of chemicals under Registration, Evaluation and Authorization of Chemicals (REACH) under European Union (EU) and usage of the 3Rs – reduction, replacement and refinement – in animal research.

In vivo testing can only be used for the determination of general toxicity on the whole body of organisms. Besides the unethical treatment of animals, there are some other troublesome aspects in *in vivo* testing like the cost of chemicals being tested, experimental setup and the disposal of additional toxic chemicals to the environment while assessing their toxicity. To overcome these problems, some alternative methods can be used. One of them is to use cell or tissue cultures instead of the whole animal to determine cytotoxicity which is called *in vitro* testing. It can effectively provide information on cytotoxicity in molecular and cellular basis, hence, it is important not only because of indicating the effects of toxicants on organism but also the cellular mechanism behind those effects (Bury et al., 2014). There are also several other advantages of *in vitro* testing. It allows a quick analysis of a large number of chemicals for toxicity and provides great test opportunities at controlled and defined environments without considering many environmental and physiological factors (Stadnicka-Michalak et al., 2014). However, to count *in vitro* testing as a reliable alternative, the correlation between *in vivo* and *in vitro* testing is of concern (Ukelis et al., 2008). There are many

studies searching for the relationship between *in vivo* and *in vitro* toxicities (Babich et al., 1986; Babich and Borenfreund, 1987; Brandão et al., 1992; Castaño et al., 1996).

In the environmental regulatory domain, there is a need to prioritize testing, reduce testing and in some cases eliminate testing entirely regarding the diversity of chemicals posing a broad range of potential toxicological concerns (Richard, 2006). Therefore, in the absence of *in vitro* and/or *in vivo* experimental data, Quantitative Structure-Activity/Toxicity Relationship (QSA/TR) models are used to predict the toxic potencies of a wide variety of xenobiotics. QSA/TR studies are based on the idea that the biological activity of a chemical can be predicted or characterized by its structure (Sullivan et al., 2014). A QSA/TR model is basically a mathematical expression of the relationship between biological activity/toxicity/physicochemical property of a molecule and its chemical structure. Such studies are useful in terms of identification, screening and prioritization of dangerous chemicals as well as providing data for a wide range of chemicals without experimental data.

1.1. Aim of the Study

The main objective of this study is to integrate *in vitro* and *in silico* approaches. To this end, at first, in order to develop valid and robust QSTR models (*in silico* approach) in line with OECD principles (OECD, 2007), three different *in vitro* endpoints for topminnow (*Poeciliopsis lucida*) hepatoma cell line (PLHC-1) relevant to the enzymatic activity, lysosomal damage and mitochondrial activity were used. Experimental cytotoxicity values reported as the induction of ethoxyresorufin-O-deethylase enzyme (EROD assay) for enzymatic activity, the uptake and accumulation of 3-amino-7-dimethylamino-2-methylphenazine hydrochloride dye (NR assay) for lysosomal activity and the cleavage of 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide in mitochondria (MTT assay) for mitochondrial activity were compiled from the literature. In line with this objective to develop valid and robust QSTR models for the same endpoint (lysosomal damage), cytotoxicity values of diverse chemicals measured with NR assay for another two different fish species, i.e., epithelial cell line of fathead minnow (FHM) and scale tissue of goldfish (*Carassius auratus*) (GFS) were also compiled from the literature.

The second objective is to estimate the cytotoxicity values for a diverse set of xenobiotics with no experimental data using the generated valid and robust *in silico* models for the five endpoints and three fish cell lines. The third objective is to report the predictive coverage of the generated *in silico* models for the external set chemicals, in this way, to fill the data gap where the information on lysosomal activity, metabolic impairment and xenobiotic metabolism is lacking, The fourth objective

is to evaluate the use of the fish cell line assays for the prediction of acute fish toxicity, at least for the newly emerging pollutants, i.e., pharmaceuticals and personal care products (PPCPs). The fifth objective is to screen the most and the least toxic chemicals among the external set chemicals with no fish cytoxicity data.

2. THEORETICAL BACKGROUND

2.1. Cell Lines vs. Primary Cell Cultures

Primary cells and cell lines are major components of *in vitro* methods. Primary cells are obtained directly from the cells, tissues or organs of an organism and cell lines are developed from primary cells. They are different from each other in some aspects: firstly, primary cell cultures have a limited lifespan and are slow in proliferation, while cell lines are easy to be cryopreserved and proliferate at relatively higher rates. This feature of cell lines provides reproducibility of experiments leading more valid results although it is not always possible to have identical samples for primary cell culture in a different context, e.g. at another time, using a different individual from the same species or the same individual at a different physiological status etc. (Bols et al., 2005). Moreover, cell lines are better in the sense that they are homogeneous samples of the target organ but primary cell cultures are heterogeneous which means that they need to be separated to have a representative sample of the target organ. In this study, data set used to build QSTR models included toxicity values for three different cell lines. They are given in Table 3.1.

Table 2.1. Dataset of different fish cell lines used for QSTR modeling.

Name of the culture	Fish Species	Target organs and/or cell lines
FHM	*Pimephales promelas* (Fathead minnow)	Permanent line of epithelial cells (Gravell and Malsberger, 1965)
GFS	*Carassius auratus* (Goldfish)	Scale tissue *(*Akimoto et al., 2000*)*
PLHC-1	*Poeciliopsis lucida* (Topminnow)	Hepatocellular carcinoma; liver; hepatoma cell line *(*Babich et al., 1991; Bols et al., 2005*)*

2.2. Cytotoxicity

Cytotoxicity can be defined as toxic potencies of chemicals measured through cell lines. It can be measured with either general/basal toxicity or injury to specific cells and their functions (Bols et al., 2005). The former that refers to impairment to cellular activities shared by all or most cells is being widely used in different ways called as cell viability assays. Depending on target cell or biomarker, these assays grouped into various categories such as cell membrane integrity, lysosomal damage, metabolic impairment, cell detachment etc. Experimental cytotoxicity data compiled from literature in the present study were based on lysosomal damage and metabolic impairment.

2.2.1. Lysosomal damage

Lysosomal damage is measured by the uptake and accumulation of 3-amino-7-dimethylamino-2-methylphenazine hydrochloride, Neutral Red (NR) dye, in the lysosome of cells. The molecular structure of NR dye is given in Figure 2.1. The dye is a weak cationic molecule which forms hydrophobic bonds with radicals in lysosomes of cells. Its penetration into a cell membrane occurs with diffusion after which it accumulates in lysosomes (Repetto et al., 2008). The dye is then extracted from the lysosomes and its absorbance was measured spectrophotometrically. The absorbance is proportional to the number of living cells (Borenfreund et al., 1988).

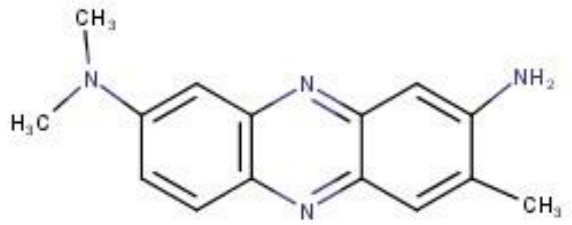


Figure 2.1. Two-dimensional molecular structure of 3-amino-7-dimethylamino-2-methylphenazine hydrochloride, Neutral Red (NR) dye. (Structure was drawn using MarvinSketch 16.2.29).

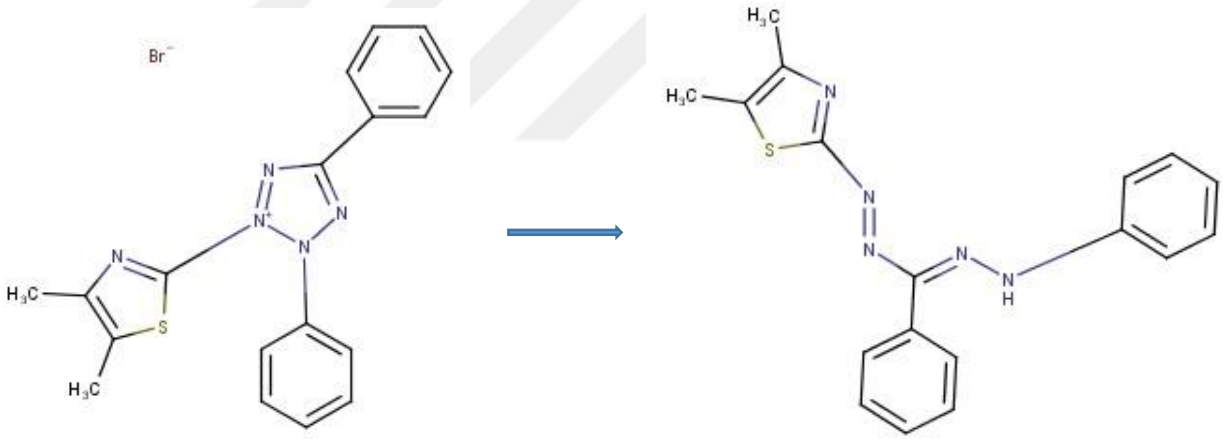
The assay was developed by Borenfreund and Shopsis (1985) which is basically based on the fact that only living cells are able to retain the dye, whereas nonliving cells cannot and impaired cells have deficiencies in uptake (Caminada et al., 2006). This results from either accumulation or retention of the dye depending on the pre- or post-exposure to toxicants (Bols et al., 2005). The endpoint value obtained from the NR assay refers to the concentration of toxicant needed for the uptake of the NR dye by 50%.

NR assay have been frequently used for a variety of purposes such as to determine cytotoxicity of metals to bluegill (BF-2) cells by Babich and co-workers (1986), of diverse chemicals to BALB/c mouse 3T3 fibroblast cell line by Borenfreund and co-workers, (1988). Besides, Nathalie and co-workers (2006) used NR assay to assess phototoxic hazard potency of certain constituents of personal care products and Asensio and co-workers (2007) used to determine the effects of metal exposure on two different earthworm species, *Eisenia fetida* and *Lumbricus terrestris*).

2.2.2. Metabolic impairment

Metabolic impairment in a target cell can be observed through the mitochondrial activity which can be measured by the cleavage of certain dyes in mitochondria. One of these dyes is 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT). Its structure is shown in Figure 2.2.(a).

MTT assay is a quantitative colorimetric assay that is used to measure cell survival and activation (Mosmann, 1983). MTT salt is cleaved by metabolically active cells in mitochondria producing blue formazan crystals (Figure 2.2.(b)) which can be measured spectrophotometrically (Bols et al., 2005). Nonliving and biologically inactive cells cannot convert the dye into formazan so that living and active cells can be detected. The endpoint value obtained from the MTT assay refers to the concentration of toxicant needed to reduce the cleavage of the MTT dye by 50%. This can be determined spectrophotometrically as the amount of formazan produced.



(a) (b)

Figure 2.2. Two-dimensional molecular structures of (a) MTT dye (b) MTT formazan product. (Structures were drawn using MarvinSketch 16.2.29).

MTT assay has been used in a variety of ecotoxicology studies. Fotakis and Timbrell (2006) used it for the determination of cadmium chloride toxicity to two hepatoma cell lines, HTC and HepG2. Borenfreund and co-workers (1988) used the MTT assay for the cytotoxicity of a range of chemicals to BALB/c mouse 3T3 fibroblast cell line. Moreover, it is also widely used in screening of cells and drugs. For example, human leukemic cell lines have been screened through MTT assay by Marks and co-workers (1992) and Van de Loosdrecht and co-workers (1994).

Cytotoxicity values measured with NR and MTT assays are stated as correlating with each other by several researches (Brüschweiler et al., 1995). Also, Caminada and co-workers (2006) reported that the experimental cytotoxicity results with NR and MTT assays had nearly the same values for pharmaceuticals. Furthermore, good correlations between *in vitro* and *in vivo* cytotoxicity measured with NR and MTT assays stated for a broad range of organic and inorganic chemicals and phenols by Brandao and co-workers (1992) and Fent and Hunn (1996).

In the literature, there are only a few studies modelling cytotoxicity endpoints obtained through NR and/or MTT assays. Most of the studies were based on the relationship of *n*-octanol-water partition coefficient (log K_{ow}) with cytotoxicity. Saito and co-workers (1993) have reported a good relationship between log K_{ow} and cytotoxicity of diverse chemicals including alcohols, aromatics, phenols and pesticides to goldfish measured through NR assay. Another good correlation was indicated by Brüschwelier and co-workers (1995) between cytotoxicity measured via MTT assay for organotin compounds and log K_{ow}. Fent and Hunn (1996) found a strong correlation of log K_{ow} with NR and MTT cytotoxicity of substituted phenols, sulphonic acids and organotin compounds to topminnow. Structural heterogeneity is an important aspect of QSA/TR models, thus compounds from different classes should be addressed covering a broad range of applicability domain, which is one of the aims, in the present study.

2.3. Enzymatic Activity

In toxicological analyses, biomarkers are used to assess the impact of environmental contaminants and as indicators responding to particular toxicants. Biomarkers are measured in whole organisms, usually to evaluate exposure. However, fish cell lines are also convenient for studying xenobiotic metabolism and enzymes and can also be used conveniently to investigate the molecular and cellular mechanisms responsible for the biomarker. One of the biomarkers used for fish is EROD activity. The most important enzyme in xenobiotic metabolism is CYP1A or P4501A, which is a member of the cytochrome P450 mixed-function oxidase system. The induction of CYP1A in fish has been evaluated as a sensitive, convenient, “early warning” signal of xenobiotics in the aquatic environment.

Induction of CYP1A enzyme is a specific response of cells when they are exposed to certain chemicals including polychlorinated biphenyls (PCBs), polychlorinated dibenzodioxins (PCDDs), dibenzofurans, and certain polyaromatic hydrocarbons (PAHs). Induction of CYP1A is a useful

biomarker as the enzyme is found at very low levels in fish liver under uncontaminated environment while its concentration increases with exposure to pollutants (Bols et al., 2005).

Induction has been demonstrated by measuring CYP1A catalytic activity as aryl hydrocarbon hydroxylase (AHH) or 7-ethoxyresorufin O-deethylase (EROD) activities. Metabolic pathway of CYP1A induction in a cell starts with the activation of aryl-hydrocarbon (AhR) which is a ligand-activated receptor. After receptor-ligand (i.e., xenobiotic) binding, AhR mediates several enzymatic events and induction of proteins (Figure 2.3 (a)).

As a biomarker in fish, the induction of CYP1A is measured through catalytic Ethoxyresorufin-O-deethylase (EROD) activity (Figure 2.3. (b)). In other words, induction is monitored as an increase in EROD activity. The amount of induced CYP1A enzyme is calculated as an increase in resorufin concentration (Figure 2.3. (b)) (mol/mg/min) and reported as the concentration of toxicant needed to increase the concentration of resorufin by 50%.

Although certain chemicals may inhibit EROD induction/activity, this interference is generally not a drawback to the use of EROD induction as a biomarker. Measurement of EROD, whose molecular structure is given in Figure 2.3. (b), as a biomarker for CYP1A activity has been widely used in fish studies (Bartram et al., 2012 and references therein).

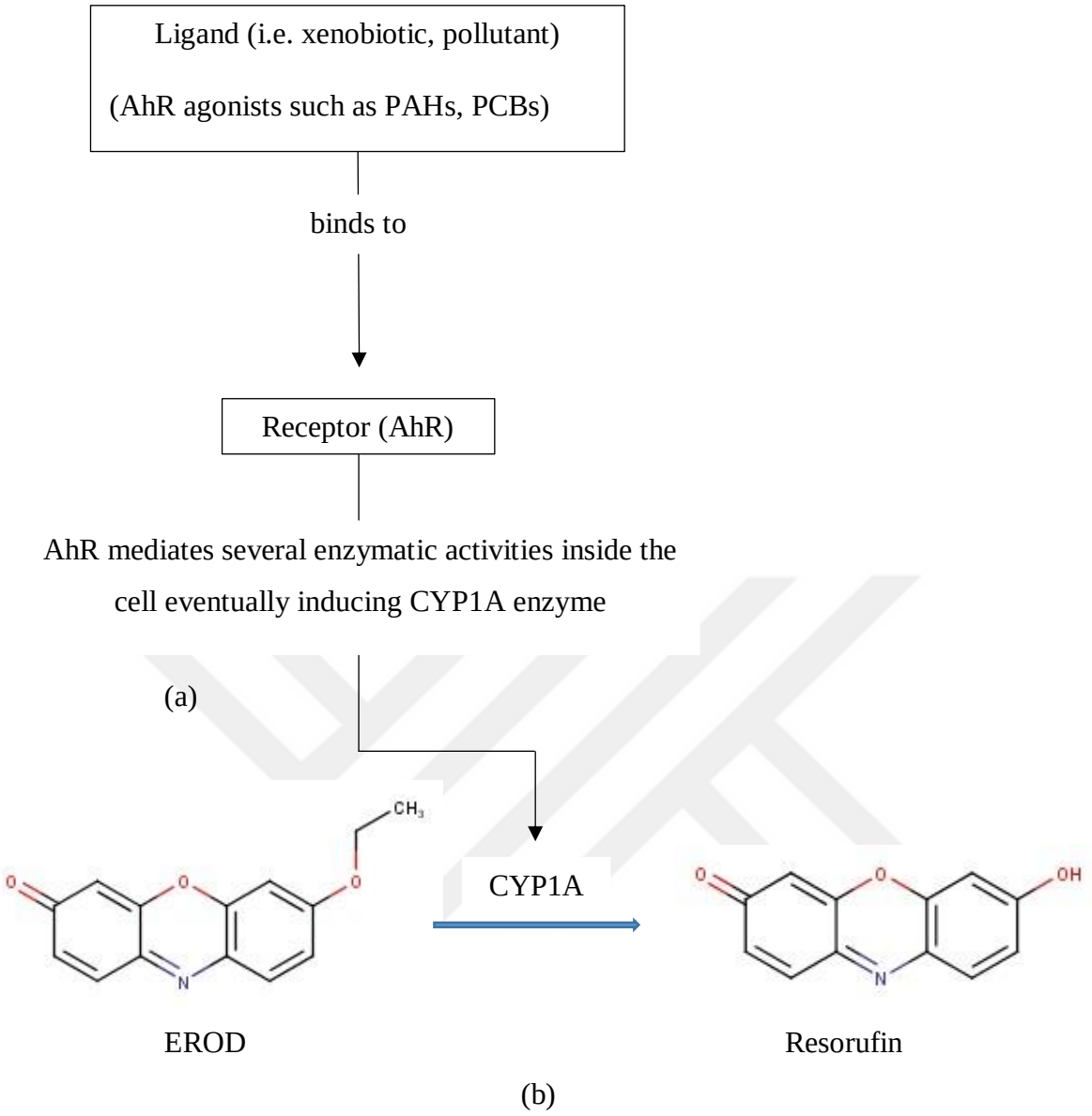


Figure 2.3. Schematic representation of the induction of CYP1A enzyme; (a) AhR mediated induction (b) EROD induction. (Structures were drawn using MarvinSketch 16.2.29).

2.4. Polycyclic Aromatic Hydrocarbons (PAHs)

Polycyclic aromatic hydrocarbons (PAHs) are chemicals that are naturally found in coal, crude, and refined mineral oils. Combustion of such materials is a major reason for the release of PAHs into the environment and they can easily be transported as particles bind to air molecules. As most of the PAHs are highly carcinogenic and mutagenic, their environmental occurrence and potential effects on organisms are of concern (Balch et al., 1995). The chemical structure of one of the most familiar PAHs, naphthalene and another PAH, perylene are shown in Figure 2.4. Some environmentally widespread pollutants like nitrated polycyclic aromatic hydrocarbons (NPAHs) and n-heterocyclic aromatic hydrocarbons (azaarenes) are derivatives of PAHs.

(a) (b)

Figure 2.4. Chemical structures of (a) naphthalene and (b) perylene (Structures were drawn using MarvinSketch 16.2.29).

2.4.1. Nitrated polycyclic aromatic hydrocarbons (NPAHs)

Nitrated polycyclic aromatic hydrocarbons (NPAHs) are derivatives of PAHs with an addition of one or more nitro group(s) to the aromatic ring. Similar to their parent compounds NPAHs are potential carcinogen and mutagens and they have also acute and chronic toxic effects (Bandowe and Meusel, 2017). Even though NPAHs are found in the environment in lower amounts compared to PAHs, their presence in various environments such as air, soil, marine, and freshwater has been detected (Jung et al., 2001). Moreover, their toxic potencies are reported as higher than those of PAHs (Collins et al., 1998). NPAHs disperse environment with incomplete combustion of fossil fuels and biomass together with PAHs. Atmospheric NPAHs generated in this way may also be deposited into the soil and aquatic environment. Due to their hydrophobic character, NPAHs are of importance considering their bioaccumulation potential.

2.4.2. Azaarenes

Azaarenes are heterocyclic compounds including a nitrogen in place of a carbon atom in the cycle (Figure 2.5). Azaarenes are carcinogenic, mutagenic and toxic compounds just as PAHs and NPAHs and they are also prevalent in the atmosphere through which they reach terrestrial and aquatic systems. Unlike PAHs and NPAHs, which are relatively more lipophilic with high *n*-octanol-water partition coefficient values (log K_{ow}), azaarenes have much lower log K_{ow} values leading to lower mobility hence lower aquatic toxicity. However, despite their low lipophilicity azaarenes can still penetrate into biological membranes and have a bioaccumulation capacity (Pearlman et al., 1984).

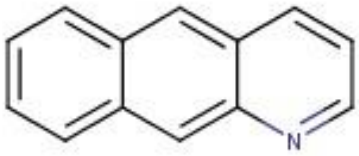


Figure 2.5. Chemical structure of acridine. (Structure was drawn using MarvinSketch 16.2.29).

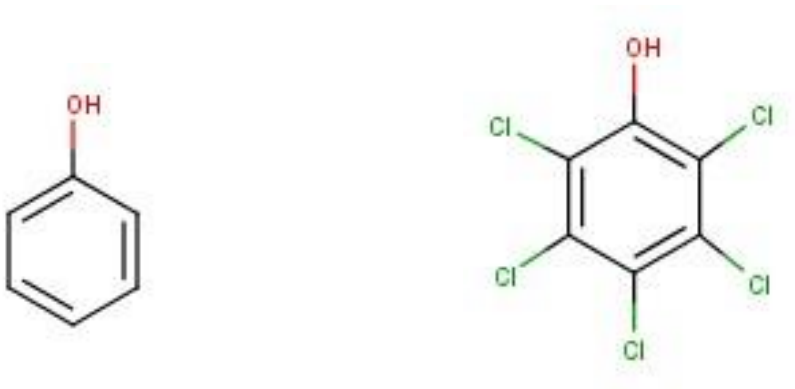
2.5. Phenols

Phenols are compounds that include a hydroxyl group bonded to an aromatic ring. They constitute one of the most important groups among aromatic compounds. They are used for a wide range of industrial applications such as dye, polymer and drug production. Phenols are used for pesticide production resulting in the environmental occurrence due to the usage of those pesticides.

Toxic potencies of phenols are related with two features: first, hydrophobicity that leads to compounds' penetration into cell membrane such that they can interact with the cell units and second, its ability to form free radicals inside the cells leading the ultimate toxic effects. Information on source and health concerns related to most commonly occurring phenols are discussed in the following parts.

Phenol (C_6H_5OH, Figure 2.6. (a)), the monohydroxy derivative of benzene, is one of the compounds included in The List of Priority Pollutants by the US Environmental Protection Agency (US EPA, 2016). Phenol is often used for industrial purposes like the production of chemicals, fuel processes etc. and it is also a component of various pesticides, dyes and textile goods (Michałowicz and Duda, 2007). Moreover, its usage as a laboratory reagent also contributes to the environmental occurrence of phenol. Aquatic concentration of phenol ranges from 0.01 – 2.0 μg/L in natural waters (Michałowicz and Duda, 2004) to 40 mg/L in freshwater polluted via disposal of sewage from petrol industry (Bruce et al., 1987).

Acute effects of phenol are characterized by malfunctions in the skin, lungs, liver, kidney and also urinal system. It can easily penetrate into the skin with a short-term contact and even lethal effects might arise. Chronic effects such as sleeping disorders, undesired weight loss, weakness, headache are reported as a result of exposure to phenols (IARC, 1972).



(a) (b)

Figure 2.6. Chemical structures of (a) phenol (b) pentachlorophenol. (Structures were drawn using MarvinSketch 16.2.29).

Chlorophenol constitutes the largest and environmentally widespread group of phenols. The main source of chlorophenols is mono and polyaromatic compounds in soil and water forming chlorophenols as a result of chlorination. Therefore, it can be claimed that drinking water has also some amount of chlorophenols due to organic content and chlorine atoms substituted with the purpose of disinfection (Michalowicz and Duda, 2006).

The presence of chlorophenols in the environment are mainly driven by industrial use. They are commonly used as pesticides and textile preservatives and impurities from the raw materials used in the production of dyes. The highest aquatic concentration of one of the most important chloropnenol derivative, pentachlorophenol (Figure 2.7. (b)), was detected as much as 0.1-10 mg/L in industrial effluents. The oral toxicity (LC_{50}) was reported as 146-175 mg/kg for adult Sherman strain rats (Gaines, 1969). Additionally, in a study by Schweigert and co-workers (2001), the carcinogenic effects of chlorophenols on mammals were also reported.

Similar to chlorophenols and phenol, the environmental occurrence of nitrophenols result from both natural and anthropogenic process. The natural generation of nitrophenols can be exemplified with their production through the reaction of phenol with nitrite ions in natural waters (Michałowicz and Duda, 2007). The anthropogenic sources, on the other hand, include pesticides by which nitrophenols are formed during production as well as after degradation. Other industrial processes like polymer, dye, solvent, plastic and drug production also lead to generation/usage of nitrophenols, as well.

2.6. Pesticides

Pesticides are chemicals used to kill or control undesired organisms like insects, pests, fungi etc. They are used to protect crops so that they meet the world's food needs and sustain the economy by increasing food production. Moreover, some pests are potent threats to public health besides depleting crops. However, even though pesticides have beneficial effects as they provide preservation of crops, their environmental effects are of concern regarding the exposure of the human and other organisms to them. For example, pesticides classified as broad-spectrum agents are not only toxic to pests but they can also be toxic to nontarget organisms.

For environmental risk assessment, it is important to consider the persistence and environmental fate of pesticides to assess how and to what extent an acute or chronic effect occurs. As a result of aerial or ground spraying, pesticides might be emitted into the air so that they can reach human, wildlife, and soil. Indeed, once applied, pesticides can disperse to all environmental compartments (e.g., aquatic, atmospheric environments etc.) through environmental processes like deposition, runoff, and leaching. In this way, pesticides may also reach drinking water posing a great health concern. Some pesticides are volatile that can lead to inhalation of dangerous particulate materials. Moreover, pesticides, such as aldrin, 1,1'-(2,2,2-Trichloroethane-1,1-diyl)bis(4-chlorobenzene) (DDT), endrin etc. are very persistent in the environment thus they have the potential of biomagnification and long-term effects.

Some acute effects of pesticides including allergic sensitization and irritation can be exemplified with asthma, dermatitis and lacrimation for former and respiratory tract irritation for the latter (Sanborn et al., 2002). Acute pesticide exposure also characterized by gastrointestinal problems like diarrhea and urination (Sanborn et al., 2002)

2.7. Active Pharmaceutical Ingredients (APIs)

Active Pharmaceutical Ingredients (APIs), a group of chemicals used for the treatment or prevention of illnesses and for veterinary health care, are emerging environmental contaminants. Considering health care systems, the pharmaceutical industry takes up a very important part of human and veterinary medicine. Both synthetic and natural substances, that are pharmaceutically active, are present in the market. As the name implies, natural substances are originated from plant and animal sources whereas industrial techniques, involving microbiological and chemical processes, are

necessary for the generation/production of synthetic drugs. APIs can be categorized depending on the therapeutic effects as gastrointestinal agents, analgesics, respiratory agents, antineoplastics etc.

Although usage of drugs is under administrative control by certain prescriptions and dosage restrictions, their environmental fate is of significance from a risk assessment perspective. The occurrence of APIs in the environment and their potential ecological effects are of concern because they are continually disposed to the environment via municipal and livestock waste streams. Various APIs have been detected in sewage effluents, surface waters, groundwaters and also surface waters (Caminada et al., 2006).

2.8. Alcohols, Aldehydes, Ketones and Esters

Alcohols are intermediate products formed in biological systems as a result of oxidation of hydrocarbons. They are classified depending on the where the hydroxyl group is attached on the molecule as primary, secondary and tertiary alcohols. Under aerobic conditions, primary and secondary alcohols can be oxidized by microorganisms producing ketones and aldehydes as intermediates.

Commercial use of alcohols includes the synthesis of organic compounds, the production of beverages and synthetic solvents. Methyl, ethyl, isopropyl and n-butyl alcohol are examples of alcohols used for commercial purposes. For example, methyl alcohol (i.e., methanol) is used as automobile antifreeze.

Aldehydes are formed by the oxidation of primary alcohols, reduction of carboxylic acids and ozonation of unsaturated hydrocarbons. Unsaturated hydrocarbons, occurring in the air as a result of automobile exhausts and incomplete combustion of fossil fuels, react with ozone. In this reaction, aldehydes play an important role in the complex system of photochemical atmospheric reactions. Aldehydes are one of the air pollutants responsible for the eye irritation.

Aldehydes like formaldehyde, propionaldehyde, acetaldehyde are used in synthesis of organic compounds and some other industrial applications. However, aldehydes are very toxic for microorganisms, if their amount in wastewater is above the toxic threshold concentration, biological treatment may not be possible.

Ketones are oxidation products of secondary alcohols. They include a carbonyl group attached to two alkyl groups. For industrial purposes, ketones are used as a solvent and for the synthesis of certain products.

Esters are products of the reaction between acids and alcohols. Most of the esters have pleasant odors, thus they are used in perfumes and as flavouring agents. They are also commercially used as solvents to separate and purify antibiotics.

2.9. Acids

Organic acids are characterized by carboxyl group(s) in a molecule. Acids that include one carboxyl group are called as monocarboxylic acid. In the case of including more than one carboxyl group they are called as polycarboxylic acid.

Monocarboxylic acids are naturally found in fats, oils and waxes. Some other monocarboxylic acids, oleic, linoleic and linolenic acids also present in many fats and oils. Microorganisms use organic acids as a nutrient which is limited by the water solubility of a molecule. A polycarboxylic acid, adipic acid, is widely used in nylon industry leading to its occurrence in industrial wastes. Other examples include oxalic acid, malonic acid and glutaric acid.

2.10. Quantitative Structure-Activity/Toxicity Relationships (QSA/TRs)

Quantitative Structure-Activity/Toxicity Relationships (QSA/TR) basically aims at deriving models to predict/explain biological activity/toxicity of a compound from its molecular structure. Such dependence of biological activity/toxicity of a compound on its structure can be encoded by physicochemical properties, 2D and/or 3D molecular properties, geometric, topological, chemical properties etc. The steps for the generation of a QSA/TR model is summarized in Figure 2.7.

A QSA/TR model is simply a mathematical function providing a measure of target activity, which is called as “endpoint”, through a measure of chemical structure (OECD, 2007). To transform the chemical structure of a compound into a modeling variable, certain mathematical and logical approaches are defined such that it can be expressed as a number, i.e. “molecular descriptor” (Todeschini and Consonni, 2008).

It can be claimed that modern QSA/TR methodology has arisen with the independent studies of Hansch and Fujita, and Free and Wilson in 1964. Both studies provided a novel and developed approaches in QSA/TR methodology later called as Hansch analysis and Free-Wilson analysis, respectively.

To build QSA/TR models, certain regression and classification methods are used. Of these methods, Multiple Linear Regression (MLR) based on ordinary least squares (OLS) method was used in the present study. MLR is a statistical approach used to build models that derive a linear equation from the relationship of a dependent variable with a number of independent variables.

Organization for Economic Co-operation and Development (OECD) states that a QSA/TR model should have the following properties to be reliable and acceptable (OECD, 2007):

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness, and predictivity;
5. a mechanistic interpretation, if possible.

An endpoint, which could be any biological, physicochemical or environmental property, refers to what is to be modeled in a QSA/TR study. For a valid QSA/TR model, there should be a defined endpoint with data that is determined by the same standardized and verified test protocols.

The methodology, which is followed in a QSA/TR study, should be clearly defined. It should be provided in detail according to OECD principle 2: an unambiguous algorithm. For a valid QSA/TR model, steps followed should be transparent. Therefore, dataset with endpoint and all used molecular descriptor values, software and/or methods used for the derivation of descriptors, training/test set divisions, resulting mathematical equation and also statistical parameters for model validation should be reported.

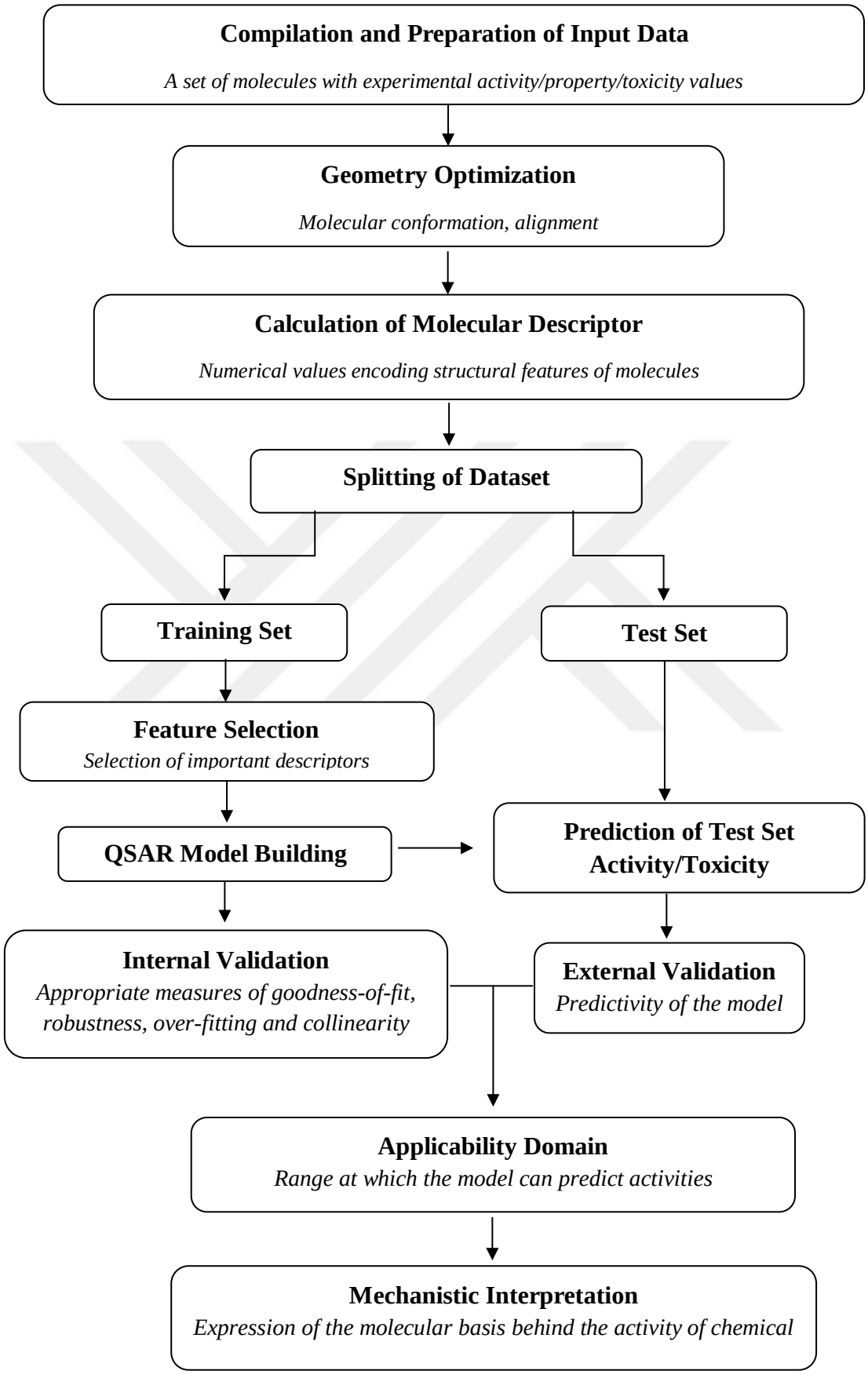


Figure 2.7. Basic steps of a QSA/TR model.

Applicability Domain (AD) of a model is a useful tool to determine/state the limits of a model in predicting an endpoint reliably. Hence, it is important to provide an AD for a QSA/TR model. In the present study, the AD of each QSTR model was defined with Williams plot which is based on the leverage approach. Standardized residuals plotted against the hat value of each chemical forms the Williams plot.

A QSA/TR model should be assessed in terms of appropriate measures of goodness of fit and robustness (internal validation) and in terms of predictivity (external validation). For statistical validation, various parameters are addressed such as the coefficient of determination (R^2), Leave-One-Out cross-validation correlation coefficient (Q^2_{LOO}), Concordance Correlation Coefficient (*CCC*) etc. These parameters were discussed in the material and method section in detail.

Finally, OECD proposes a mechanical interpretation for QSA/TR models, if possible. Mechanicalinterpretation refers to the explanation of modeling endpoint in the context of model descriptors hence to point out the processes/reasons behind the activity/toxicity of compounds (OECD, 2007).

3. MATERIALS AND METHODS

3.1. Datasets

Datasets compiled from the literature on cytotoxicity and EROD activity of various chemicals from different assays and fish species/cell lines are listed in Table 4.1 and used for QSTR modeling. A detailed description of each dataset was given in the following sections. In the present study, three endpoint values used for QSTR modeling were expressed as $pEC_{50,NR}$ which refers to the negative logarithm of the concentration of toxicant needed to reduce of the uptake of the NR dye by 50%, $pEC_{50,MTT}$ which refers to the negative logarithm of the concentration of toxicant needed to reduce the cleavage of the MTT Dye by 50% and $pEC_{50,EROD}$ which refers to the negative logarithm of concentration of toxicant needed to induce the concentration of resorufin by 50%.

Table 3.1. Experimental data compiled from the literature.

Toxicity Endpoint	Chemicals	Fish Species/Cell Line	Reference
$EC_{50,EROD}$	5 compounds (4 azoles and β-naphthoflavone)	Topminnow *Poeciliopsis lucida* (Hepatoma cell line) PLHC-1	Gräns et al., 2015
	11 compounds (PAHs)		Fent and Bätscher, 2000
	3 compounds (TCDD, 2 PCBs)		Hahn et al., 1996
	23 compounds (PAH, NPAH and azaarenes)		Jung et al., 2001
$EC_{50,NR}$	16 human pharmaceuticals	Topminnow *Poeciliopsis lucida* (Hepatoma cell line) PLHC-1	Caminada et al. 2006
	35 organic chemicals	Fathead minnow *Pimephales promelas* Epithelial cell line FHM	Brandáo et al., 1992

Table 3.1. Continued.

$EC_{50,NR}$	100 chemicals of different classes	Goldfish *Carassius auratus* Scale tissue GFS	Saito et al., 1993
	17 pesticides	Topminnow *Poeciliopsis lucida* (Hepatoma cell line) PLHC-1	Knauer et al., 2007
	26 organic compounds		Fent and Hunn, 1996
$EC_{50,MTT}$	21 human pharmaceuticals	Topminnow *Poeciliopsis lucida* (Hepatoma cell line) PLHC-1	Caminada et al. 2006
	23 organic compounds		Fent and Hunn, 1996
	8 pharmaceuticals		Laville et al., 2004

3.1.1. Dataset for $EC_{50,EROD}$ Endpoint

The negative logarithm of the concentration of toxicant needed to induce the concentration of resorufin by 50% measured for *Poeciliopsis lucida* hepatoma cell line is named as $pEC_{50,EROD[PLHC-1]}$. The $pEC_{50,EROD}$ data from Grans and co-workers (2015) including for 5 compounds, 4 azoles and β-naphthoflavone, were not included as their laboratory procedure did not comply with the procedures of the other three studies (Table 3.1). Additionally, $pEC_{50,EROD}$ values for some of the compounds were reported both by Fent and Bätscher (2000) and Jung and co-workers (2001). In this case, the average values of experimental $pEC_{50,EROD}$ values were taken for the common chemicals.

Fent and Bätscher (2000) studied the EROD induction potency of 11 PAHs and reported this induction as $EC_{50,EROD}$ in molar [M]. Similarly, the EROD induction potency of 23 compounds including PAHs, NPAHs and azaarenes reported as $EC_{50,EROD}$ [M] by Jung and co-workers (2001). Finally, Hahn and co-workers (1996) provided $EC_{50,EROD}$ values for TCDD, two planar PCBs in [nM]. All of the $EC_{50,EROD}$ values were first converted into mM, then transformed to the negative logarithmic scale and named as $pEC_{50,EROD[PLHC-1]}$. Ultimate dataset for $pEC_{50,EROD[PLHC-1]}$ composed of 32 diverse chemicals with experimental $pEC_{50,EROD[PLHC-1]}$ values ranging from 0.444 to 6.886 [mM].

3.1.2. Dataset for $EC_{50,NR}$ Endpoint

The concentration of toxicant needed to reduce the uptake of the NR dye by 50% is designated as $EC_{50,\ NR}$. The $EC_{50,NR}$ dataset was somehow more complicated than either $EC_{50,EROD}$ or $EC_{50,MTT}$

datasets. For this endpoint experimental data for three different fish species/cell lines, namely, *Poeciliopsis lucida* hepatoma cell line (PLHC-1), *Pimephales promelas* epithelial cell line (FHM) and *Carassius auratus* scale tissue (GFS) were available. Therefore, the $EC_{50,\ NR}$ values for the three different fish species are named as $EC_{50,NR[PLHC-1]}$, $EC_{50,NR[FHM]}$ and $EC_{50,NR[GFS]}$, referring to each cell line (Table 3.2).

Table 3.2. Experimental $EC_{50,NR}$ data compiled from the literature for the three fish cell lines.

$EC_{50,NR}$	Cell Line	Reference
$EC_{50,NR[PLHC-1]}$	PLHC-1	Fent and Hunn, 1996 Knauer et al., 2007 Caminada et al. 2006
$EC_{50,NR[GFS]}$	GFS	Saito et al., 1993
$EC_{50,NR[FHM]}$	FHM	Brandáo et al., 1992

3.1.2.1. Dataset for the $EC_{50,NR[PLHC-1]}$. This dataset comprises experimental $EC_{50,NR}$ values of 57 diverse chemicals. $EC_{50,NR}$ values of; 25 organic chemicals including substituted phenols, sulphonic acids and organotin compounds as $EC_{50,NR}$ [M] were taken from Fent and Hunn (1996), 12 pesticides as $pEC_{50,NR}$ [M] were taken from Knauer and co-workers (2007), and 20 pharmaceuticals as $pEC_{50,NR}$ [mM] were taken from the study of Caminada and co-workers (2006). All $EC_{50,NR}$ values are converted to mM and then transformed to the negative logarithmic scale and named as $pEC_{50,NR[PLHC-1]}$. The range of the transformed dataset is from -2.80 to 2.66 [mM].

3.1.2.2. Dataset for $EC_{50,NR[GFS]}$. The experimental data $EC_{50,NR}$ measured for goldfish, *Carassius auratus,* scale tissue (GFS) were taken from the study of Saito and co-workers (1993). This dataset comprises $EC_{50,NR,[GFS]}$ values [mM] of chemicals from various classes, such as alcohols, aromatics, anilines, phenols, aldehydes and pesticides. The negative logarithm of $EC_{50,NR,[GFS]}$ [mM] values is named as $pEC_{50,NR[GFS]}$. The range of $pEC_{50,NR[GFS]}$ values is from -3.02 to 2.96.

3.1.2.3. Dataset for $EC_{50,NR[FHM]}$. The experimental data $EC_{50,NR}$ measured for fathead minnow *Pimephales promelas,* epithelial cell line (FHM) were taken the study of Brandáo and co-workers (1992). The $EC_{50,NR[FHM]}$ values for a variety of organic chemicals, such as alcohols, esters, ketones and inorganic chemicals were provided as mM in their work. Experimental results for only organic compounds were compiled and the negative logarithm of $EC_{50,NR[FHM]}$ [mM] named as $pEC_{50,NR[FHM]}$. The final data set included 33 chemicals with $pEC_{50,NR[FHM]}$ values ranging from -2.96 to 0.11.

3.1.3. Dataset for $EC_{50,MTT}$ Endpoint

The $EC_{50,MTT}$ dataset comprised 27 pharmaceuticals (Caminada et al., 2006; Laville et al., 2004) and 23 organic chemicals (Fent and Hunn, 1996). The reported cytotoxicity endpoint was $EC_{50,MTT}$ for fish (*Poeciliopsis lucida*) hepatoma cell line, PLHC-1. 8 pharmaceuticals whose cytotoxicity values adopted from the work of Laville and co-workers (2004) were not used in QSTR modeling, due to differences in experimental procedure. Dataset composed of 42 chemicals. $EC_{50,MTT}$ values of 19 pharmaceuticals were taken from the study of Caminada and co-workers (2006) (reported as mM) and 23 organic chemicals including substituted phenols, sulphonic acids and organotin compounds were taken from Fent and Hunn, (1996) (reported as M). All of the cytotoxicity values are converted to mM and transformed to negative logarithm scale and named as $pEC_{50,MTT[PLHC-1]}$. Experimental $pEC_{50,MTT[PLHC-1]}$ values were between -2.48 and 2.59.

3.2. Calculation of Molecular Descriptors

For the generation of a QSTR model the steps which are summarized in Figure 2.8 has been followed. In the first step, the molecular structure of compounds was sketched and geometrically optimized with the semi-empirical PM6 method using SPARTAN 10 (Wave-function, 2010). The lowest energy conformations of molecular geometries were selected for the molecular descriptor calculations.

Geometrically optimized molecular structures were saved as mol2 files and then loaded into DRAGON 06 (Talete, 2013) software for the descriptor generation. The DRAGON software generates descriptors from 20 different blocks, i.e. atomic-molecular; physicochemical; electronic; geometric properties of the molecules (e.g., topological indices, connectivity indices, 3D-MoRSE etc.)

Additionally, the SPARTAN software also generates some descriptors: energy of the lowest unoccupied molecular orbital (E_{LUMO}), the energy of the highest occupied molecular orbital (E_{HOMO}), the energy in the aqueous phase (E_{aq}), dipole moments, gas–phase energy (E), CPK volume and area (V and A, respectively). Moreover, some other descriptors: $E_{LUMO} - E_{HOMO}$ gap, hardness (η), electronegativity, softness (σ), and electrophilicity index (ω) were calculated using these descriptors from SPARTAN 10 with formulas reported by LoPachin and co-workers (2007). A pool of about 5000 descriptors was initially uploaded to QSARINS software. QSARINS 2.2.1. software (Gramatica et al. 2013, 2014; QSARINS 2015) was used for the selection of descriptors and model development.

SPARTAN is a useful tool to calculate most stable conformation of molecules and to make geometry optimization but it might not be useful for all chemicals. In this study, problems for some of the compounds have arisen during the calculation of conformer distribution. Therefore, for compounds whose molecular conformation calculation was not possible, Molecular Mechanics (MMF) method embedded in SPARTAN 10 was used. Additionally, SPARTAN 10 software could not calculate aqueous phase energy (E_{aq}, whose value is used to determine the lowest energy conformation of a molecule) for some chemicals. In this case, the gas phase energy of the compounds was used to choose the lowest energy conformer.

3.3. *In silico* Modeling

QSTR/*in silico* models for different endpoints were generated using QSARINS 2.2.1. software in accordance with OECD principles. As OECD principle 4 implies, generated models were checked in terms of both internal performance and external predictivity. Therefore, initial datasets were divided into training (~80% of compounds) and test sets (~20% of compounds) in order to develop and verify the predictive capability of the proposed models, respectively. Different division methods have been applied: (i) by the ordered response, (ii) by ordering the molecules based on their structure, and (iii) random divisions.

After dividing datasets into training and test sets, MLR models were generated using All subset and Genetic Algorithm (GA)-based iterative facilities implemented in the QSARINS 2.2.1 software (Gramatica, 2016). All subset selection simply covers all the descriptor combinations that can be formed from a descriptor pool reaching a user-defined number of modeling variables. GA tool in QSARINS 2.2.1 is useful in terms of providing a chance to modify population size, the mutation rate and the number of generations for the genetic algorithm.

3.4. Internal Validation Parameters

QSTR models generated in the present study were checked in terms of robustness with the square of determination coefficient (R^2), the square of determination coefficient adjusted for the number of variables ($R^2_{adj.}$) and Fischer statistics (F). Root Mean Square Errors (RMSE) were assessed for training and test sets, $RMSE_{TR}$ and $RMSE_{TEST}$, respectively.

3.4.1. R^2 (Coefficient of Determination)

R^2 is the coefficient of determination that basically expresses variance on the information provided by a regression model (Rao, 1973 as referred in Nagalkerke, 1991). It is a statistical tool to understand how well a regression predicts a dependent variable from independent variable(s). The general formula for R^2 is the following equation (Eq. 3.1):

$$R^2=1-\frac{\Sigma\left(Y_{\text{obs}}-Y_{\text{pred}}\right)^2}{\left(Y_{\text{obs}}-\overline{Y_{\text{obs}}}\right)^2} \tag{3.1}$$

where $Y_{observed}$ and $Y_{calculated}$ are observed and predicted response values (i.e. dependent variable) while $\overline{Y}_{obs}$ is the mean value of all the observed response values. R^2 can take values between 0 and 1 with an ideal model having R^2 = 1 whereas, for R^2 > 0.5 a regression model is acceptable.

3.4.2. R^2_{adj} (Adjusted R^2)

The value of R^2 keeps increasing by addition of new variables to the regression model. However, additional variables do not always contribute to the model in predicting the dependent variable but points out a statistical development. To overcome this, a new parameter R^2_{adj}, in which the number of degrees of freedom is also considered besides variance in observed and predicted response values, has been created.

In addition to R^2_{adj}, another criterion was also used to avoid an excess number of variables in a model. The ratio of the number of regression data to the number of modeling variables is limited to 5 that is called as Topliss ratio (Topliss and Costello, 1972).

3.4.3. F (Variance Ratio)

Fischer statistics (F) value provides information about the statistical importance of a regression model. Higher F value indicates a more significant model and it is calculated by the ratio of explained and unexplained variance in data for a given degree of freedom. It is calculated with the following formula (Eq. 3.2):

$$F=\frac{\frac{\Sigma(Y_{\text{pred}}-\overline{Y})^2}{p}}{\frac{\Sigma(Y_{\text{obs}}-Y_{\text{pred}})^2}{N-p-1}} \tag{3.2}$$

where N is the number of experimental response values and p is the number of independent variables used to predict the response values. Y_{pred} is the predicted and Y_{obs} is the experimental response values and whereas $\overline{Y}$ is the mean experimental response value.

3.4.5. *Y*-scrambling

Chance correlations were controlled with Y-scrambling response randomization tests by which different models were generated by a random response. A significantly low coefficient of determination the new models (for R^2_{Yscr} and Q^2_{Yscr} new coefficients of determination following *Y*-scrambling procedure) indicate that there is no correlation by chance. The Y-scrambling procedure was run in QSARINS 2.2.1.

3.4.6. Leave-One-Out Cross-Validation Correlation Coefficicent (Q^2_{LOO})

To test internal predictivity, leave-one-out cross-validation correlation coefficient (Q^2_{LOO}) was used; in this parameter, one randomly chosen compound was excluded from the training set and the toxicity of that compound was calculated using the originally selected descriptors. In this way, the respective importance of individual variables in the model could be seen. To calculate Q^2_{LOO} value Eq. 3.3 is used.

$$Q^2_{LOO}=1-\frac{\Sigma(Y_{obs}-Y_{pred})^2}{\Sigma(Y_{obs(TR)}-\overline{Y}_{TR})^2} \quad (3.3)$$

where Y_{obs} and Y_{pred} are experimental and leave-one-out predicted response values, respectively, $Y_{obs(TR)}$ is the experimental response values for the training set and $\overline{Y}_{TR}$ is the mean experimental response value for the training set compounds.

3.5. External Validation Parameters

3.5.1. Predictive Squared Correlation Coefficients (Q^2_{F1}, Q^2_{F2}, Q^2_{F3})

Q^2_{F1} shows correlation between predicted and observed toxicity values (Shi et al., 2001) (Eq. 3.4):

$$Q^2_{F1}=1-\frac{\Sigma(Y_{obs(TEST)}-Y_{pred(TEST)})^2}{\Sigma(Y_{obs(TEST)}-\overline{Y}_{TR})^2} \quad (3.4)$$

where $Y_{obs(TEST)}$ and $Y_{pred(TEST)}$ are experimental and predicted response values for the test set, respectively and $\overline{Y}_{TR}$ is the mean experimental response value for the training set compounds.

Q^2_{F2} is based on prediction of test set compounds (Schüürmann et al. 2008) (Eq. 3.5):

$$Q^2_{F2}=1-\frac{\Sigma(Y_{obs(TEST)}-Y_{pred(TEST)})^2}{\Sigma(Y_{obs(TEST)}-\overline{Y}_{TEST})^2} \quad (3.5)$$

The only difference between Q^2_{F1} and Q^2_{F2} parameters is that mean experimental response value for the training set compounds used in calculation of Q^2_{F1} is replaced with mean experimental response value for the test set compounds in Q^2_{F2}.

Finally Q^2_{F3} parameter developed by Consonni and co-workers (2009, 2010) (Eq. 3.6):

$$Q^2_{F3}=1-\frac{[\Sigma(Y_{obs(TEST)}-Y_{pred(TEST)})]^2/n_{TEST}}{[\Sigma(Y_{obs(TEST)}-\overline{Y}_{TR})]^2/n_{TR}} \quad (3.6)$$

where n_{TR} and n_{TEST} are number of compounds in training and test sets, respectively. The threshold values for these parameters were determined as Q^2_{F1}, Q^2_{F2}, Q^2_{F3}=0.70 (Chirico and Gramatica, 2012).

4.5.2. Concordance Correlation Coefficient (*CCC*)

The Concordance Correlation Coefficient **(***CCC*) parameter developed by Lin (1989; 1992) is calculated both for training and test set data (Chirico and Gramatica, 2011; 2012). For the test set it is calculated with the equation below (Eq. 3.8):

$$CCC_{TEST}=\frac{2\Sigma(Y_{obs(TEST)}-\overline{Y}_{obs(TEST)})(Y_{pred(TEST)}-\overline{Y}_{pred(TEST)})}{\Sigma_{i=1}^{n}(Y_{obs(TEST)}-Y_{obs(TEST)})^2+\Sigma_{i=1}^{n}(Y_{pred(TEST)}-\overline{Y}_{pred(TEST)})^2+n(Y_{obs(TEST)}-\overline{Y}_{pred(TEST)})} \quad (3.8)$$

where $Y_{obs(TEST)}$ and $Y_{pred(TEST)}$ are experimental and predicted response values for the test set compounds while $Y_{obs(TEST)}$ and $\overline{Y}_{pred(TEST)}$ refer to the average values of the experimental and predicted response values for the test set compounds. In an ideal model, *CCC* has the value of 1

however, it generally takes lower values and the threshold value for CCC_{TEST} is 0.85 (Chirico and Gramatica, 2012).

3.5.3. r_m^2 Metrics

The r_m^2 parameter depends on the correlation between experimental and predicted response values. For calculation, squared correlation coefficients from two regression lines first experimental values are plotted on Y-axis and predicted values on X-axis and second predicted values are plotted on y-axis and experimental values on x-axis (Ojha et al., 2011). In an ideal model (i.e. a model with perfect predictions), these two lines intercept and correlation between two are the same. On the other hand, if there is no intercept then above situation is not valid and the two lines have different correlation coefficients.

The r_m^2 parameter basically compares the correlation between experimental and predicted response values when there is an intercept and when there is not (Ojha et al., 2011). It can be calculated with the following formula (Eq. 3.9):

$$r_m^2 = r^2\left(1-\sqrt{r^2-r_0^2}\right) \tag{3.9}$$

where r^2 and r_0^2 are squared correlation coefficients between experimental and predicted response values with and without an intercept, respectively. The threshold value for r_m^2 is 0.5 (Ojha et al., 2011).

3.5.4. Golbraikh and Tropsha Criteria

The criteria developed by Golbraikh and Tropsha (2002) was used to test the external predictive ability of the models. The criteria require following conditions to count a model as acceptable:

$R_{CV}^2 > 0.5$,

$R^2 > 0.6$,

R_0^2 or $R_0'^2$ close to R^2

i.e.: **(a)** $(R^2-R_0^2)/R^2 < 0.1$ and $0.85 \leq k \leq 1.15$ or

(b) $(R^2-R_0'^2)/R^2 < 0.1$ and $0.85 \leq k' \leq 1.15$

$\left|R_0^2-R_0'^2\right| < 0.3$ (3.10)

where R^2 is predicted vs. observed; R'^2 is observed vs. predicted; k and k' are slopes; R_0^2 and $R_0'^2$ are squared correlation coefficients.

3.5.5. Mean Absolute Error (MAE)-Based Criteria

Another criterion to test the models' performance on external prediction was developed by Roy and co-workers (2016). It depends upon Mean Absolute Error (MAE) value which is used to determine prediction errors in generated models and calculated as follows (Eq. 3.11):

$$MAE = \frac{1}{n} \times |Y_{\text{obs}} - Y_{\text{pred}}| \tag{3.11}$$

where n is the number of compounds, $Y_{obs.}$ and $Y_{pred.}$ are experimental and predicted endpoint values.

According to MAE-based criteria, a good model should fulfill the following criteria to be an acceptable model.

$$MAE \leq 0.1 \times \text{training set range} + 3 \times \sigma \leq 0.2 \times \text{training set range and}$$
$$MAE \leq 0.1 \times MAE + 3 \times \sigma \leq 0.2 \times \text{training set range}$$

where σ represents the standard deviation of the absolute error values for the test set data.

In addition to parameters mentioned here, to prevent over-fitting caused by the high number of descriptors, the ratio of the number of compounds used in training sets to the number of model variables limited to Topliss ratio (Topliss and Costello, 1972). To prevent collinearity among variables, QUIK rule which determines the difference between total correlation among the descriptors K_{XX} and the correlation among the descriptor and responses K_{XY} was used. The model is excluded if $K_{XY} - K_{XX} < \delta_K$ where the threshold value δ_K is set to 0.05 (Todeschini et al., 1999).

3.6. Applicability Domain

The applicability domain (AD) of models was generated with the leverage approach (Tropsha et al., 2003) and standardized residuals. Chemicals out of the standardized residual ($\pm 3\sigma$) are stated as response outliers. The critical hat value (h^*) is determined by $h^* = 3(p+1)/n$ (n is the number of training compounds, p is the number of descriptors in the model) (Golbraikh and Tropsha, 2002). The

compounds with leverage values higher than h^* are stated as structural outliers. The ranges of descriptors that are to be appeared in the model, toxicity values, and the leverage approach all together form the AD of a model.

3.7. Selection of the Best Model

In order to select the best model, at first the generated models were ranked based on their Multicriteria Decision Making (MCDM) (Keller et al. 1991) criteria evaluated by QSARINS 2.2.1 sooftware. MCDM summarizes the performances of a definite number of criteria relevant to internal and external validations as scores between 0 and 1 (where 0 represents the worst validation criteria value and 1 the best), simultaneously. The model with the highest MCDM score was selected as the best one (QSARINS 2.2.1, 2015).

In addition, the best models were judged by means of the fulfillment of the all validation requirements of OECD (OECD 2007). The final models were used to predict the cytotoxicity (measured with NR and MTT assays) and EROD activity) of an external set of chemicals with no experimental data.

3.8. External Set

Predictive ability of the models generated in this study was further evaluated with an additional set of compounds (i.e. external set) with no experimental data of the target endpoint. Initially, a large external set composed of 660 chemicals from various classes were used for all the models. The composition of this dataset is given in Figure 3.1.

A single QSA/TR model is not expected to predict activity of all the chemicals present however, there is a range in between the model could predict the target activity. Whatever we can obtain/expect from a model mostly depends on the dataset that the whole study based upon. Therefore, it would be meaningless to expect a QSA/TR model to predict activity of chemicals that have extremely different structures from the modelling data set chemicals. Considering this, external set for the models generated for $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[FHM]}$ endpoints were revised and restricted to the compounds with at least some structural similarity.

The external set for the $pEC_{50,EROD[PLHC-1]}$ model composed of 353 chemicals and its composition is given in Figure 4.2. On the other hand, Figure 4.3 provides composition of external set for $pEC_{50,NR[FHM]}$ model which is composed of 230 chemicals.

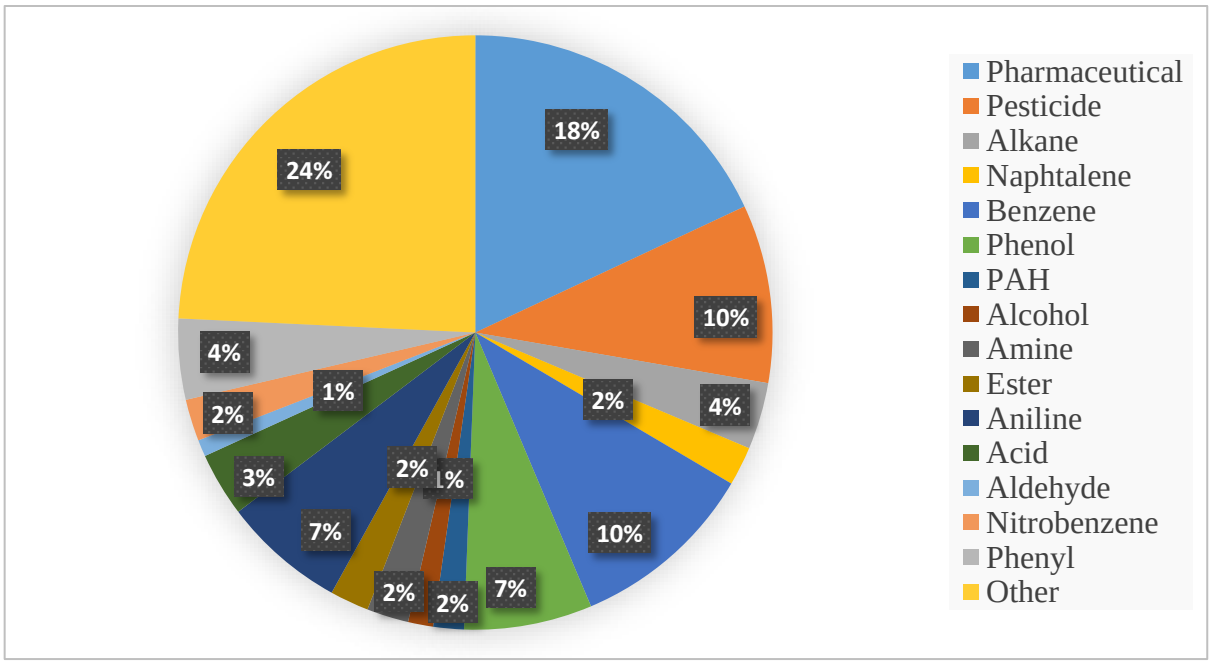


Figure 3.1. Chemical classes of 660 compounds in the external set of $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ models.

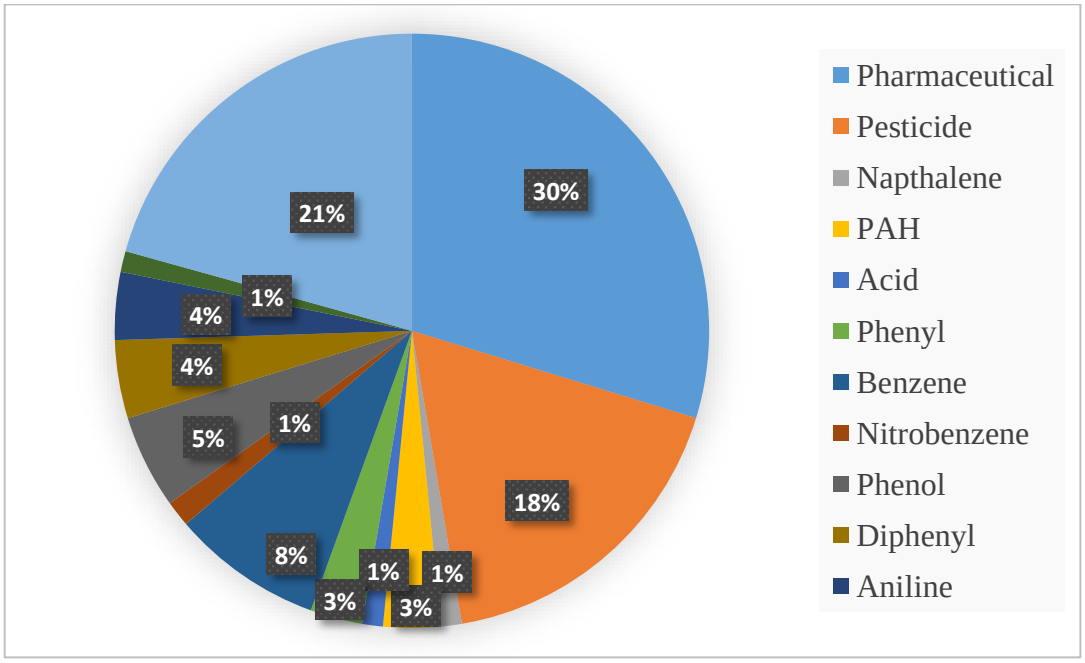


Figure 3.2. Chemical classes of 353 compounds in the external set of $pEC_{50,EROD[PLHC-1]}$ model.

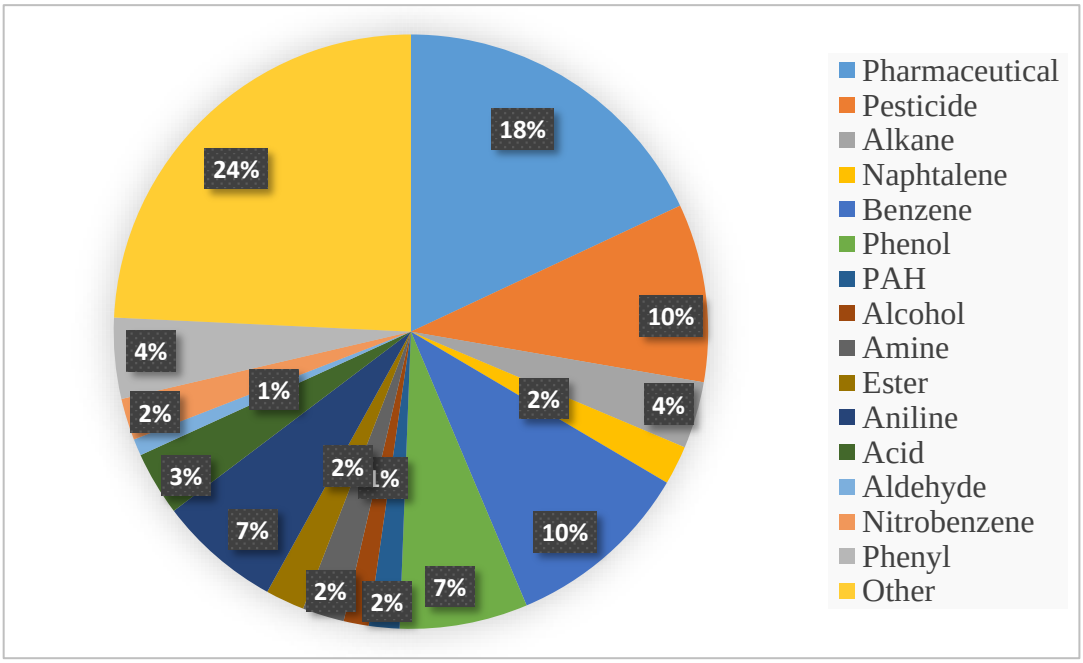


Figure 3.3. Chemical classes of 230 compounds in the external set of $pEC_{50,EROD[FHM]}$ model.

3.9. Correlation Analyses

The correlation analyses were performed between (i) log K_{ow} and experimental cytotoxicity values of each endpoint (ii) the predicted cytotoxicity values from each of the $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$, $pEC_{50,NR[FHM]}$ and $pEC_{50,MTT[PLHC-1]}$ models, and (iii) the predicted cytotoxicity values (*in vitro*) and fish lethality (*in vivo*) data. All the reported analyses were carried out in SPSS 24 (IBM SPSS, 2016) using Pearson correlation coefficient (*r*) when the two-tailed significance level was set to 0.01.

The correlation analysis between the predicted cytotoxicity values for each endpoint and fish lethality data was implemented to assess the usability of *in vitro* predictions as an alternative to *in vivo* toxicity testing. The *in vivo* data as 96-h fish lethal concentration (LC_{50}) of 230 external set chemicals reported for *Oryzias latipes* were obtained from the aquatic toxicity database of the Japan Ministry of the Environment. In this database, lethal concentrations were reported as mg/L, they were first converted to mM and then to the negative logarithm scale and named as pLC_{50} in the present study.

4. RESULTS AND DISCUSSION

4.1. QSTR Modelling for $pEC_{50,EROD[PLHC-1]}$

4.1.1. Dataset

Chemicals in the dataset with their physicochemical properties and experimental EROD induction potency values ($pEC_{50,EROD[PLHC-1]}$ as mM) are listed in Table 4.1. Experimental $pEC_{50,EROD[PLHC-1]}$ values of five chemicals were reported by both Fent and Bätscher (2000) and Jung and co-workers (2001) therefore, their average values were taken and used in modeling.

Table 4.1. The name, CAS number, log K_{ow}, molecular weight and experimental $pEC_{50,EROD[PLHC-1]}$ values of the chemicals in the dataset.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,EROD[PLHC-1]}$ (mM)	References
1	1,6-dinitropyrene	42397-64-8	4.42	292.250	2.796	Jung et al., 2001
2	1-azapyrene	313-80-4	3.43	203.244	1.824	Jung et al., 2001
3	2-azafluoranthene	7148-92-7	3.01	203.244	1.745	Jung et al., 2001
4	2-nitrofluorene	607-57-8	3.80	211.220	1.824	Jung et al., 2001
5	2-nitronaphthalene	581-89-5	3.07	173.171	1.071	Jung et al., 2001
6	3,3',4,4',5-pentachlorobiphenyl	57465-28-8	6.50	326.437	6.432	Hahn et al., 1996
7	3,3',4,4'-tetrachlorobiphenyl	32598-13-3	5.94	291.992	4.108	Hahn et al., 1996
8	3-methylcholanthrene	56-49-5	1.72	268.359	4.102	Fent and Bätscher, 2000
9	3-nitrofluoranthene	892-21-7	4.38	247.253	2.022	Jung et al., 2001
10	6-Nitrochrysene	7496-02-8	5.06	273.291	3.301	Jung et al., 2001
11	7,12-dimethylbenz[a] anthracene	57-97-6	6.00	256.348	3.032	Fent and Bätscher, 2000
12	7-azafluoranthene	206-49-5	3.43	203.244	1.432	Jung et al., 2001
13	7-nitrobenzo[a]anthracene	20268-51-3	5.06	273.291	2.420	Jung et al., 2001
14	10-azabenzo[a]pyrene	24407-49-6	4.43	253.304	3.194	Jung et al., 2001
15	Benz[a]anthracene	56-55-3	5.03	228.294	2.700***	Fent and Bätscher, 2000 and Jung et al., 2001
16	Benzo[a]acridine	225-11-6	4.54	229.282	3.097	Jung et al., 2001
17	Benzo[a]pyrene	50-32-8	5.34	252.316	3.699	Fent and Bätscher, 2000 and Jung et al., 2001
18	Benzo[c]acridine	225-51-4	4.54	229.282	2.337	Jung et al., 2001
19	Benzo[e]pyrene	192-97-2	5.34	252.316	3.602	Fent and Bätscher, 2000
20	Benzo[h]quinoline	230-27-3	3.11	179.222	0.796	Jung et al., 2001

Table 4.1. Continued.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,EROD[PLHC-1]}$ (mM)	References
21	Benzo[k]fluoranthene	207-08-9	5.34	252.316	4.481	Fent and Bätscher, 2000
22	Chrysene	218-01-9	5.03	228.294	3.544***	Fent and Bätscher, 2000 and Jung et al., 2001
23	Dibenz[a,h]anthracene	53-70-3	6.02	278.354	4.927***	Fent and Bätscher, 2000 and Jung et al., 2001
24	Dibenzo[a,c]acridine	215-62-3	5.53	279.342	3.770	Jung et al., 2001
25	Dibenzo[a,c]anthracene	215-58-7	6.02	278.354	4.180	Jung et al., 2001
26	Dibenzo[a,h]acridine	226-36-8	5.53	279.342	3.359	Jung et al., 2001
27	Dibenzo[a,i]pyrene	189-55-9	6.34	302.376	4.620	Fent and Bätscher, 2000
28	Dibenzo[a,j]acridine	224-42-0	5.53	279.342	3.301	Jung et al., 2001
29	Fluoranthene	206-44-0	4.35	202.256	0.444	Jung et al., 2001
30	Perylene	198-55-0	5.34	252.316	2.959	Fent and Bätscher, 2000
31	Pyrene	129-00-0	4.35	202.256	2.357***	Fent and Bätscher, 2000 and Jung et al., 2001
32	TCDD	1746-01-6	4.28	321.974	6.886	Hahn et al., 1996

n*-octanol-water partition coefficients calculated by SPARTAN 10; ** MW from SPARTAN 10; * Average values of experimental data from Fent and Bätscher, 2000 and Jung et al., 2001.

4.1.2. Model Development

$pEC_{50,EROD[PLHC-1]}$ dataset contains 32 chemicals (Table 4.1). The dataset was divided into various training and test set combinations using response-based, structure-based and random splitting options included in QSARINS software. For each division, 1-2 descriptor models were generated. Models resulted from randomly divided datasets were slightly better than either structure or response-based divisions. Therefore, the models were selected from these divisions. The test set compounds and the ratio of the number of test set to the training set compounds in three random divisions that provided best models are given in Table 4.2.

Table 4.2. The test set compounds for three divisions used in the QSTR modeling of $pEC_{50,EROD[PLHC-1]}$.

Division no	n_{TEST}/n_{TR}	Test Set Compounds*
1	10/22	3, 11, 12, 14, 19, 25, 26, 27, 28 31
2	10/22	3, 11, 12, 14, 16, 19, 25, 26, 27, 28, 31
3	7/25	4, 11, 12, 13, 14, 16, 26

*Compound numbers refer to the ID numbers given in Table 4.1

Fit, internal and external validation parameters of the best 1-2 descriptor models are given in Table 4.3 and 4.4, respectively. The selected models fulfilled all the fit, internal and external validation criteria, therefore they are subjected to further criteria to determine the best model.

Table 4.3. Fit and internal validation parameters of the generated $pEC_{50,EROD[PLHC-1]}$ models.

No	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
				Division 1					
1.1	ZM1Mad	0.756	0.744	0.787	0.861	62.066	0.711	0.856	0.834
				Division 2					
***2.1**	**ATS3m**	**0.727**	**0.713**	**0.834**	**0.842**	**53.121**	**0.664**	**0.924**	**0.806**
2.2	Mor05m	0.724	0.710	0.837	0.705	52.465	0.657	0.933	0.800
				Division 3					
3.1	SM4_B(m) GATS3m	0.887	0.876	0.519	0.940	85.927	0.857	0.582	0.924

*The selected model.

Table 4.4. External validation parameters of the generated $pEC_{50,EROD[PLHC-1]}$ models.

No	Descriptors	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2–r_0'^2)/r^2$	$(r^2–r_0^2)/r^2$
						Division 1							
1.1	ZM1Mad	0.875	0.355	0.870	0.870	0.951	0.934	0.820	0.064	1.007	0.982	0.011	0.001
						Division 2							
***2.1**	**ATS3m**	**0.856**	**0.380**	**0.851**	**0.850**	**0.943**	**0.914**	**0.736**	**0.139**	**0.973**	**1.014**	**0.058**	**0.004**
2.2	Mor05m	0.925	0.479	0.763	0.763	0.910	0.917	0.654	0.140	1.029	0.953	0.051	0.147
						Division 3							
3.1	SM4_B(m) GATS3m	0.944	0.292	0.913	0.869	0.964	0.941	0.918	0.007	1.073	0.928	0.001	0.001

*The selected model.

4.1.3. Comparison of Applicability Domain of $pEC_{50,EROD[PLHC-1]}$ Models

The best QSTR model was selected by comparing the applicability domain (AD) of the generated models given in Tables 4.3 and 4.4. For this purpose, each model was used to predict $pEC_{50,EROD[PLHC-1]}$ values of 353 chemicals from various classes with no experimental $pEC_{50,EROD[PLHC-1]}$ data. Number of chemicals that the model could predict $pEC_{50,EROD[PLHC-1]}$ values out of 353 were given in Table 4.5 with corresponding percent structural coverage of the models.

Table 4.5. Predictive performances of the generated $pEC_{50,EROD[PLHC-1]}$ models.

Division no	Model no	Number of chemicals within AD of the model (out of 353)	Structural coverage (%)
1	1.1	291	82.2
2	*__2.1__	__325__	__92.1__
	2.2	252	71.4
3	3.1	129	36.5

*The selected model.

Regarding all the criteria that the models were subjected, the one-descriptor model 2.1 was selected as the best model and written in bold through Tables 4.3-4.4 where its fit, internal and external validation parameters were given, respectively. Its predictive performance for external set chemicals with no experimental data was given in Table 4.5. The resulting equation for model 2.1 was given below (Eq 4.1). The numbers in parentheses show the 95% confidence intervals of the coefficients.

$$pEC_{50,EROD[PLHC-1]} = 4.908\ (\pm 1.405)\ ATS3m - 15.099\ (\pm 5.243) \quad (4.1)$$

The fit of the model was good as indicated by high R^2 (0.727), R^2_{adj} (0.713) and low $RMSE_{TR}$ (0.834) values. The parameter Q^2_{LOO} = 0.664 was acceptable and comparable with the R^2 value asserting the robustness of the model. To test the reliability of the model and whether there is a possibility of chance correlation, Y-scrambling response randomization test was applied using 2000 iterations. The resulting low R^2_{Yscr} (0.049) and Q^2_{Yscr} (-0.155) values revealed that the generated model was not obtained by chance. Although the internal parameters of the model were not superior to other models generated for $pEC_{50,EROD[PLHC-1]}$ endpoint, its external validation parameters were comparable. The predictive ability of the model was reflected by a high R^2_{TEST} (0.856) and a low $RMSE_{TEST}$ (0.380). Additionally, the model had the external validation parameters Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{TEST} higher than the corresponding literature threshold values promoting its predictive power.

Closer values of Q^2_{F1} = 0.851 and Q^2_{F2} = 0.850 reflected that the response distribution of the test set was homogeneous. For all the other external validation parameters presented in Table 4.4, the model was found to be acceptable. As a further evaluation of external performance, Golbraikh and Tropsha's criteria was also followed by which the model was considered successful satisfying all of the expected conditions. Finally, the model was subjected to MAE-based criteria developed by Roy and co-workers (2016). Regarding the MAE-based criteria, the model classified as "good" with values of *MAE* (95% of the data) = 0.290, 3σ = 0.495 and training set range = 6.442.

Chemicals used for the QSTR modelling of $pEC_{50,EROD[PLHC-1]}$ together with their training/test set status in modelling, experimental and predicted $pEC_{50,EROD[PLHC-1]}$ values, hat and descriptor values are given in Table 4.6.

Table 4.6. Chemicals used for the QSTR modelling of $pEC_{50,EROD[PLHC-1]}$, their experimental and predicted $pEC_{50,EROD[PLHC-1]}$ values, hat values and descriptor values.

Chemicals	Status	Exp. $pEC_{50,EROD[PLHC-1]}$ (mM)	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h^*=0.273)	ATS3m
1,6-dinitropyrene	Training	2.796	4.174	0.070	3.927
1-azapyrene	Training	1.824	2.157	0.071	3.516
2-azafluoranthene	Prediction	1.745	1.995	0.080	3.483
2-nitrofluorene	Training	1.824	1.608	0.106	3.404
2-nitronaphthalene	Training	1.071	0.160	0.269	3.109
3,3',4,4',5-pentachlorobiphenyl	Training	6.432	5.946	0.235	4.288
3,3',4,4'-tetrachlorobiphenyl	Training	4.108	4.861	0.116	4.067
3-methylcholanthrene	Training	4.102	3.664	0.052	3.823
3-nitrofluoranthene	Training	2.022	3.114	0.046	3.711
6-nitrochrysene	Training	3.301	3.522	0.049	3.794
7,12-dimethylbenz[a]anthracene	Prediction	3.032	3.389	0.047	3.767
7-azafluoranthene	Prediction	1.432	2.010	0.079	3.486
7-nitrobenzo[a]anthracene	Training	2.420	3.536	0.049	3.797
10-azabenzo[a]pyrene	Prediction	3.194	3.512	0.048	3.792
Benz[a]anthracene	Training	2.700	2.339	0.063	3.553
Benzo[a]acridine	Prediction	3.097	2.412	0.060	3.568
Benzo[a]pyrene	Training	3.699	3.463	0.048	3.782
Benzo[c]acridine	Training	2.337	2.432	0.059	3.572
Benzo[e]pyrene	Prediction	3.602	3.556	0.049	3.801
Benzo[h]quinoline	Training	0.796	0.827	0.181	3.245
Benzo[k]fluoranthene	Training	4.481	3.232	0.046	3.735
Chrysene	Training	3.544	2.452	0.058	3.576
Dibenz[a,h]anthracene	Training	4.927	3.630	0.051	3.816
Dibenzo[a,c]acridine	Training	3.770	3.792	0.055	3.849

Table 4.6. Continued.

Chemicals	Status	Exp. $pEC_{50,EROD[PLHC-1]}$ (mM)	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	HAT value (h^*=0.273)	ATS3m
Dibenzo[a,c]anthracene	Prediction	4.180	3.718	0.053	3.834
Dibenzo[a,h]acridine	Prediction	3.959	3.703	0.052	3.831
Dibenzo[a,i]pyrene	Prediction	4.620	4.518	0.090	3.997
Dibenzo[a,j]acridine	Training	3.301	3.684	0.052	3.827
Fluoranthene	Training	0.444	1.946	0.082	3.473
Perylene	Training	2.959	3.556	0.049	3.801
Pyrene	Prediction	2.357	2.118	0.073	3.508
TCDD	Training	6.886	5.652	0.197	4.228

The linear relationship between predicted $pEC_{50,EROD[PLHC-1]}$ values from Eq 4.1 versus experimental $pEC_{50,EROD[PLHC-1]}$ values is shown in Figure 4.1. It can be seen that the data is homogeneously aligned along the optimal line. Therefore, a good agreement of the predicted $pEC_{50,EROD[PLHC-1]}$ values with the experimental data is claimed.

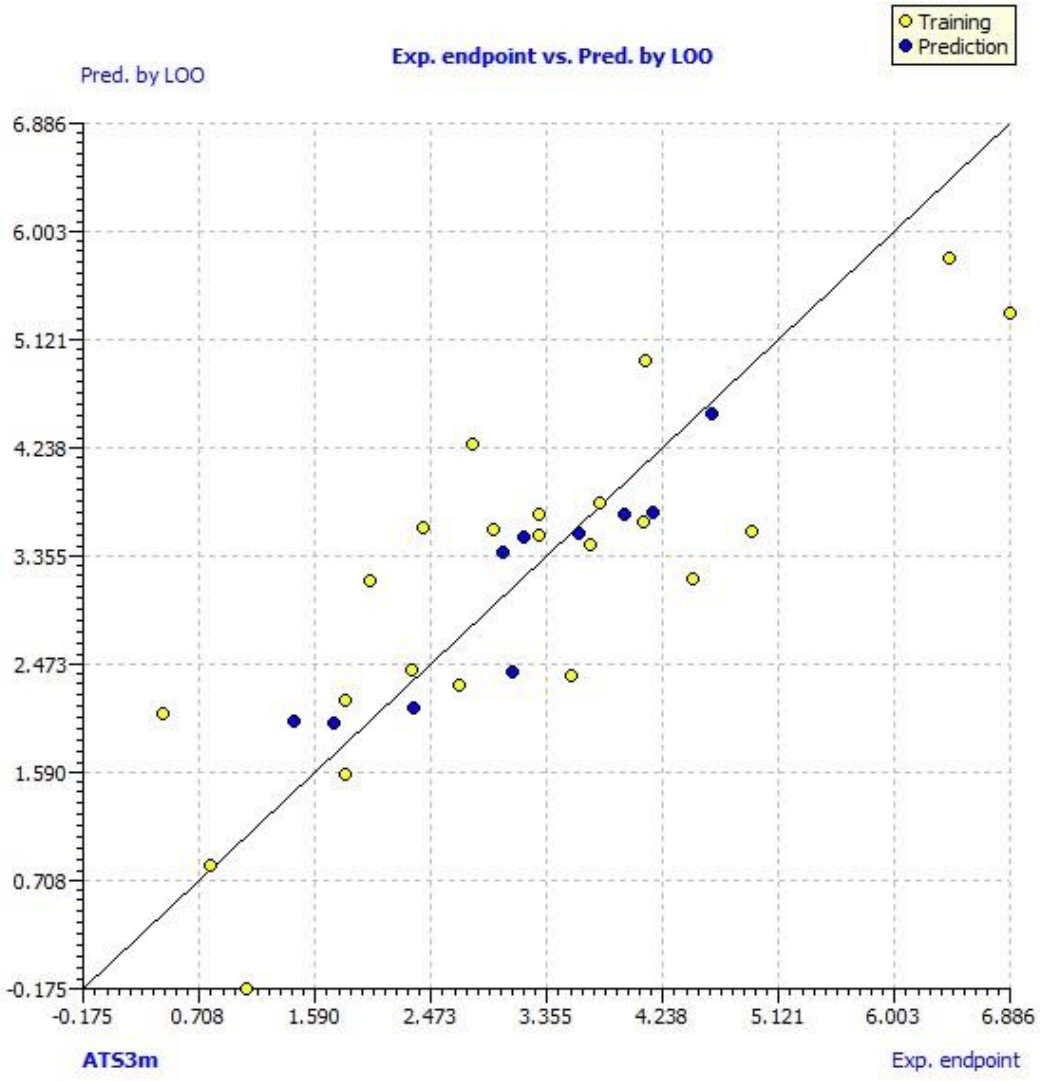


Figure 4.1. Plot of predicted $pEC_{50,EROD[PLHC-1]}$ values from Eq. 4.1 versus experimental $pEC_{50,EROD[PLHC-1]}$ values.

The plot of standardized residuals versus hat values (Williams plot) is given in Figure 4.2. With this plot the AD of the $pEC_{50,EROD[PLHC-1]}$ model can be displayed and any outlier from structural and response spaces can be detected. It is important to state that all of the chemicals in the data set had residuals lower than $\pm 2.5\sigma$ referring to the absence of a response outlier. Moreover, none of the chemicals is indicated as a structural outlier because all of them had hat values lower than the critical hat value ($h^* = 0.273$).

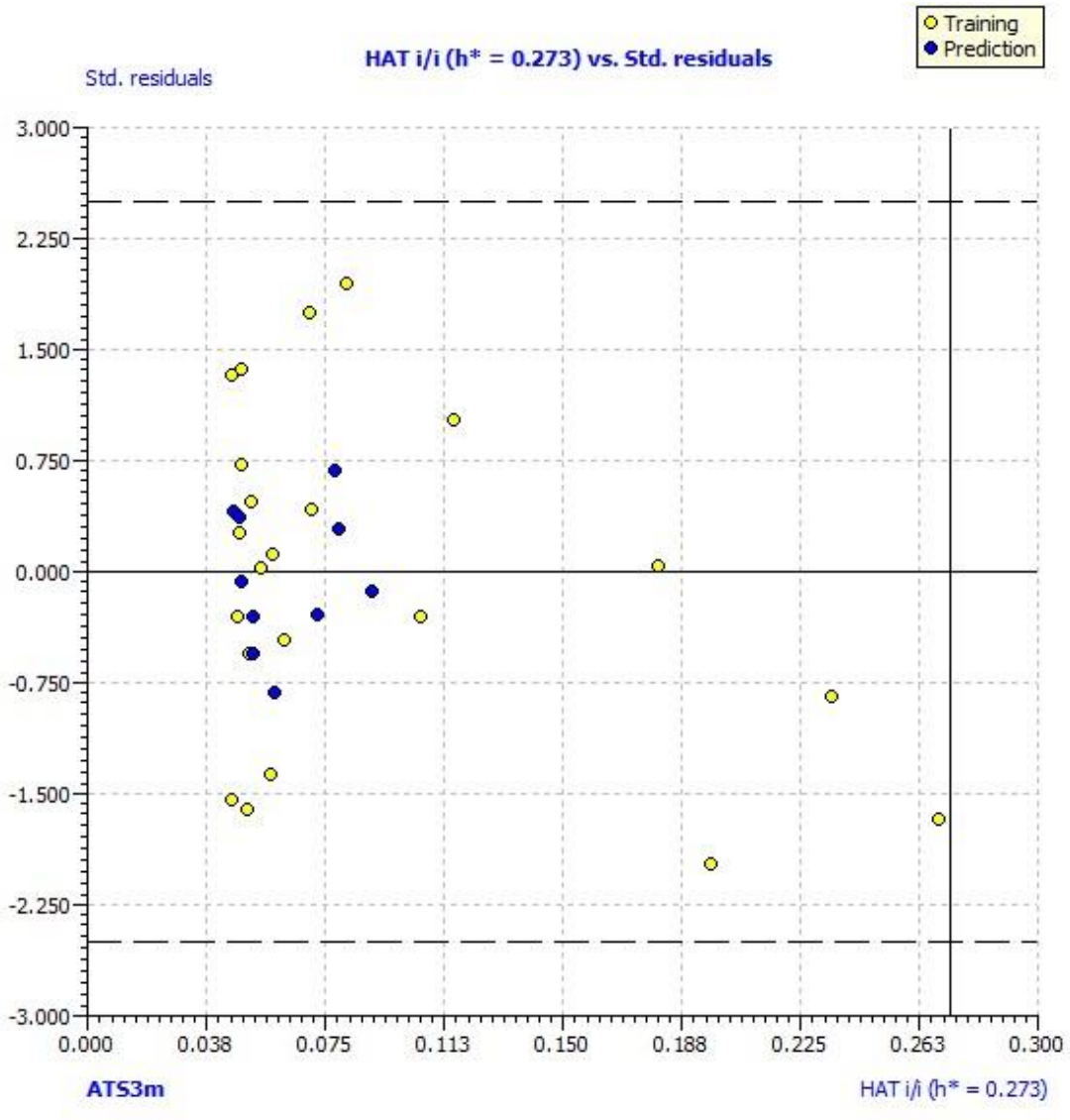


Figure 4.2. Williams plot of $pEC_{50,\ EROD[PLHC-1]}$ model.

To test the external predictive ability of the $pEC_{50,EROD[PLHC-1]}$ model, an external set composed of 353 chemicals was used. Extrenal set chemicals, calculated descriptor, predicted and experimental $pEC_{50,EROD[PLHC-1]}$ values are provided in Appendix A. Of the 353 external set chemicals, 28 chemicals were outside the AD of the generated $pEC_{50,EROD[PLHC-1]}$ model.

Insubria graph of the model showing its structural coverage is given in Figure 4.3 where the predicted $pEC_{50,EROD[PLHC-1]}$ values for training, test and external set chemicals from Eq. 4.1 versus their hat values are plotted.

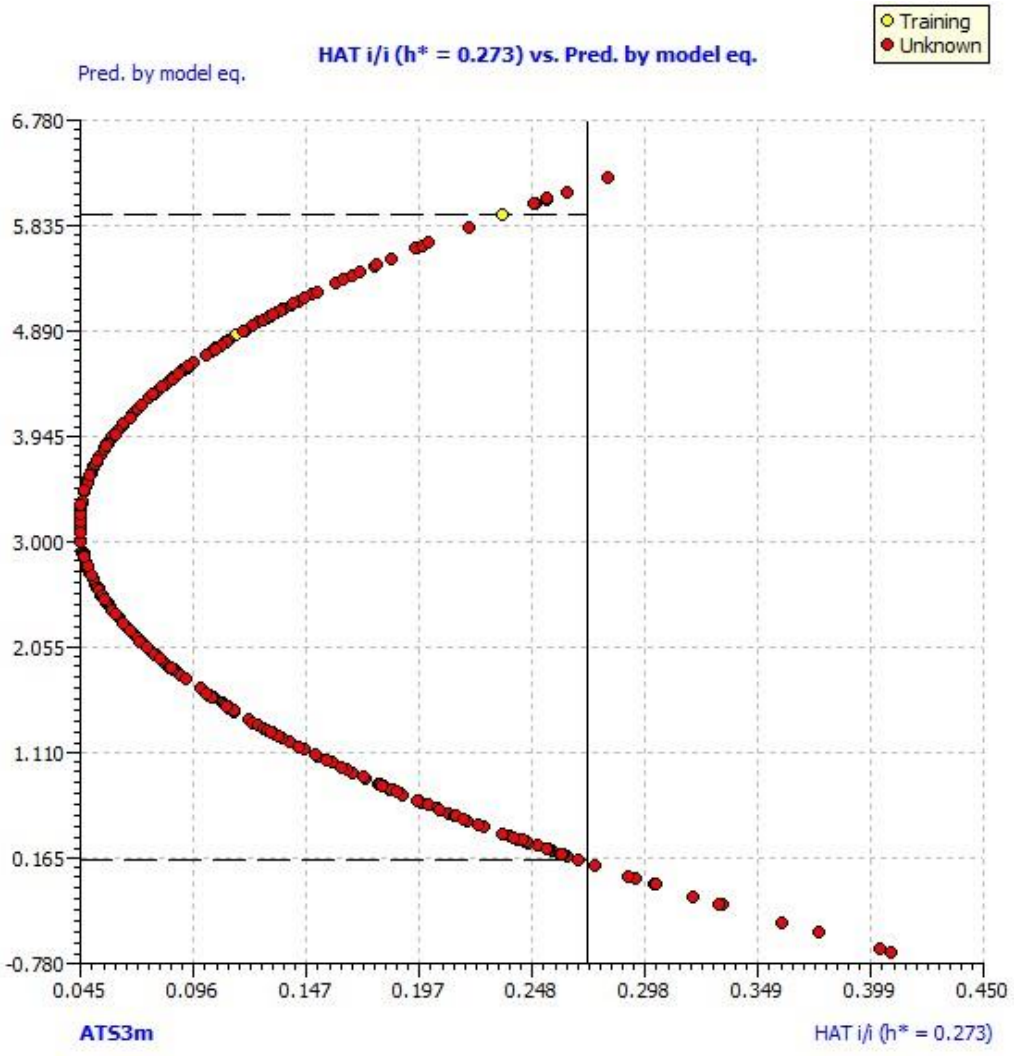


Figure 4.3. Insubria graph of $pEC_{50,EROD[PLHC-1]}$ model. Predicted $pEC_{50,EROD[PLHC-1]}$ values for training, test and external (353 chemicals) set chemicals from Eq. 4.1 versus their hat values.

Ten chemicals with the most and the least $pEC_{50,EROD[PLHC-1]}$ values predicted by Eq. 4.1 were screened and given in Table 4.7. The most toxic chemical with highest $pEC_{50,\ EROD[PLHC-1]}$ value was found as tetrachloroguaiacol.

Table 4.7. The most and the least cytotoxic chemicals from the external set predicted by Eq. 4.1.

	Name	Chemical class	$pEC_{50,EROD[PLHC-1]}$ pred. by Eq. 4.1 (mM)
Most Toxic	Tetrachloroguaiacol	Phenol, methoxy, chloro	5.833
	Tetrachlorocatechol	Benzenediol, chloro	5.686
	Testosterone propionate	Ester, propionate	5.666
	Finasteride	Pharmaceutical, Benign Prostatic Hypertrophy Agents	5.642
	Pyrazosulfuron ethyl	Pesticide, Herbicide	5.549
	2,4,6-tribromophenol	Phenol, bromo	5.539
	Camptothecin	Pharmaceutical, Antineoplastic Agents, Phytogenic	5.495
	Tetrachlorohydroquinone	Hydroquinone, chloro	5.495
	Nicosulfuron	Pesticide, Herbicide	5.485
	Methotrexate	Pharmaceutical, Antineoplastic Agents	5.426
Least Toxic	Fluorene	PAH	0.160
	Benzophenone	Diphenyl ketone	0.160
	2,4-diamino-6-nitrotoluene	Benzene, nitro, diamine	0.194
	2,6-diamino-4-nitrotoluene	Benzene, nitro, diamine	0.194
	2,4-dichlorotoluene	Benzene,chloro	0.204
	2,5-dichlorotoluene	Benzene,chloro	0.204
	2,6-dichlorotoluene	Benzene,chloro	0.204
	Cotinine	Nicotine metabolite	0.209
	2,4-diamino-6-phenyl-*s*-triazine	Phenyl-diazine,diamino	0.214
	Isopropylnaphthalene	Naphthalene	0.238

Tetrachloroguaiacol, a chlorophenolic compound, is found in effluents from the paper and pulp industry (Woodland and Maly, 1997). Rogers and co-workers (1989) reported that the concentration of tetrachloroguaiacol can reach up to 111 µg per kg of fish (*Oncorhynchus tshawytscha*) under the conditions where the fish is exposed to certain waste effluents. In addition, it has high bioaccumulation potential in aquatic organisms (Michałowicz, 2005). Therefore, further evaluation of the toxic effects of tetrachloroguaiacol would be important regarding its persistence and accumulation capacity from an environmental risk assessment perspective. The high $pEC_{50,EROD[PLHC-1]}$ value presented in this study might serve as an initial prioritization for this chemical and lead to further testing/evaluation.

The DRAGON descriptor, ATS3m, appeared in Eq 4.1 belongs to the 2D autocorrelations block and described as Broto-Moreau autocorrelation of lag 3 (log function) weighted by mass. Broto-Moreau autocorrelations are spatial autocorrelations defined as given in Eq. 4.2 (Moreau and Broto, 1980a, 1980b; Broto et al., 1984).

$$ATS_k = \frac{1}{2} \sum_{i=1}^{A} \sum_{j=1}^{A} w_i w_j \delta(d_{ij};k) \qquad (4.2)$$

where w is any atomic property like atomic mass, polarizability, charge and electronegativity; A is the number of atoms in a molecule; k is the lag; and $\delta(d_{ij};k)$ is a Kronecker delta function d_{ij} is the topological distance between i^{th} and j^{th} atoms which equals to either 1 (if $d_{ij} = k$) or zero (Todeschini and Consonni, 2009).

Autocorrelation descriptors characterize the distribution of molecular properties (such as topological or geometric distances at given topological positions). For the indicated atomic property w, ATS descriptors can be calculated for a set of topological distances expressed as $\{ATS_k\}_w$. ATS3m descriptor is calculated using the information about the separation between two atoms that are 3 lag units apart and weighted by atomic mass.

EROD induction is a result of a series of enzymatic reactions starting with the activation of aryl hydrocarbon receptor (AhR) by a xenobiotic. Therefore, ligand (xenobiotic) and protein (AhR) binding is a crucial part of the whole process. Two-dimensional molecular structure, atom-atom topological/geometric distances in a ligand (i.e. xenobiotic) might affect ligand-target protein complexation so that it could be related to the EROD induction potency of that xenobiotic. The

presence of ATS3m descriptor in the model equation can be interpreted in this way. Additionally, a weighting of ATS3m is done with mass and thus, it can be claimed that the mass of a molecule has an effect on its EROD induction potency.

Broto-Moreau autocorrelation descriptors have previously been used for the QSTR modelling of cytotoxicity of phenylindole derivatives (antimitotic agents) to two breast cancer cell lines (Adhikari et al., 2015). Also, Fereidoonnezhad and co-workers (2017) used similar descriptors to explain anti-proliferative activity of *n*-phenyl ureidobenzenesulfonate derivatives against human breast adenocarcinoma (MCF-7) cell line.

4.1.4. Correlation between $pEC_{50,EROD[PLHC-1]}$ and log K_{ow}

In order to determine whether log K_{ow} is a good predictor of EROD induction potency, the relationship between the log K_{ow} and $pEC_{50,EROD[PLHC-1]}$ values was investigated. Figure 4.4 shows the linear relationship between experimental $pEC_{50,EROD[PLHC-1]}$ values and log K_{ow} values.

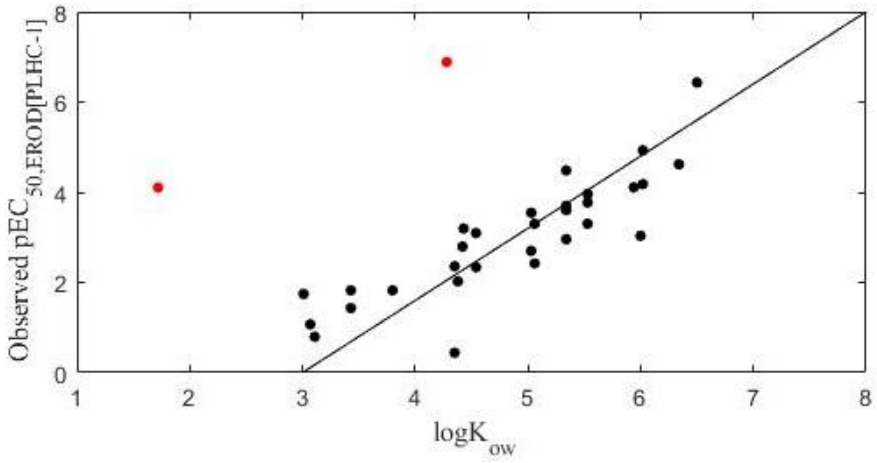


Figure 4.4. Plot of observed $pEC_{50,EROD[PLHC-1]}$ values versus log K_{ow} values.

The correlation analysis revealed that there is a strong correlation between log K_{ow} and $pEC_{50,EROD[PLHC-1]}$ values for all compounds (n = 32; r = 0.547; p < 0.01). For two chemicals 3-methylcholanthrene and TCDD, log K_{ow} and $pEC_{50,EROD[PLHC-1]}$ relationship was weak, these chemicals are indicated in Figure 4.4 with red markers. However, regarding the reported strong relationship between the log K_{ow} and $pEC_{50,EROD[PLHC-1]}$ for the remaining chemicals in the dataset, it can be claimed that log K_{ow} might be used for preliminary assessment of EROD induction potency of the studied compounds for PLHC-1 cell line.

4.1.4. Correlation of Experimental *in vivo* pLC_{50} with Predicted *in vitro* $pEC_{50,EROD[PLHC-1]}$

To assess the suitability of the proposed model for the prediction of *in vivo* toxicity, the relationship between the predicted $pEC_{50,EROD[PLHC-1]}$ values and the fish lethality (pLC_{50}) data were evaluated. The compiled pLC_{50} values from the literature for external set chemicals relevant to the studied endpoints are given in Appendix A.

A strong correlation was found between the predicted *in vitro* $pEC_{50,EROD[PLHC-1]}$ values and *in vivo* data (n = 50; r = 0.573; p < 0.01). Figure 4.5 depicts the plot of observed pLC_{50} versus $pEC_{50,EROD[PLHC-1]}$ values. Of the 71 external set chemicals, the predicted $pEC_{50,EROD[PLHC-1]}$ values of 21 chemicals have a weak relationship with the experimental pLC_{50} values therefore, they are excluded from the correlation analysis. These chemicals were mainly from chloro substituted benzene and anilines and bromo substituted phenols.

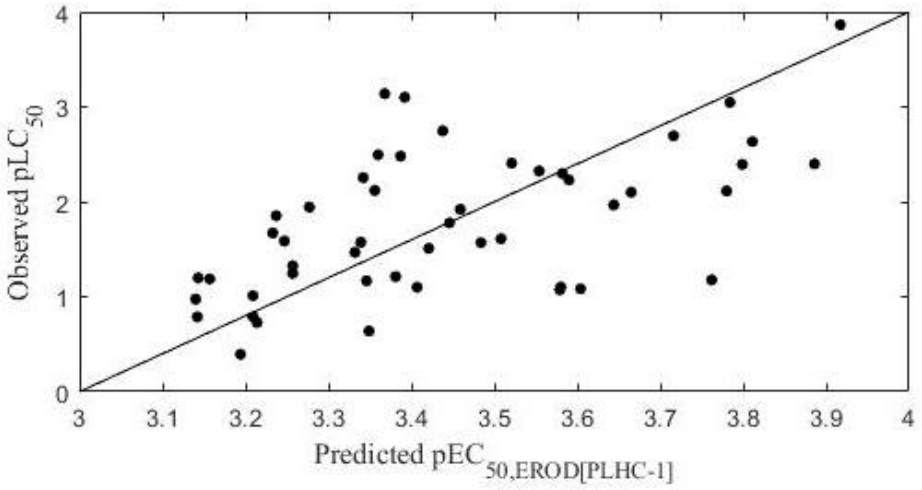


Figure 4.5. Plot of observed pLC_{50} versus predicted $pEC_{50,EROD[PLHC-1]}$ values by Eq. 4.1.

4.2. QSTR Modelling for $pEC_{50,NR[PLHC-1]}$

4.2.1. Dataset

Experimental NR-based cytotoxicity data of 57 chemicals to PLHC-1 cell line ($pEC_{50,NR[PLHC-1]}$) used in the QSTR modelling are listed in Table 4.8.

Table 4.8. The name, CAS number, log K_{ow}, molecular weight and experimental $pEC_{50,NR[PLHC-1]}$ values of chemicals in the dataset.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,NR[PLHC-1]}$ (mM)	References
1	*(S)*-propranolol	4199-09-1	0.36	259.349	0.74	Caminada et al., 2006
2	2,3,4,5-tetrachlorophenol	4901-51-3	3.88	231.893	0.77	Fent and Hunn, 1996
3	2,3,6-trichlorophenol	933-75-5	3.32	197.448	-0.18	Fent and Hunn, 1996
4	2,4,5-trichlorophenol	95-95-4	3.32	197.448	0.66	Fent and Hunn, 1996
5	2,4-dichlorophenol	120-83-2	2.76	163.003	0.55	Fent and Hunn, 1996
6	2,4-dinitrophenol	51-28-5	1.71	184.107	0.09	Fent and Hunn, 1996
7	2,5-dinitrophenol	329-71-5	1.71	184.107	0.28	Fent and Hunn, 1996
8	2,6-dichlorophenol	87-65-0	2.76	163.003	-0.72	Fent and Hunn, 1996
9	2-chlorophenol	95-57-8	2.20	128.558	0.08	Fent and Hunn, 1996
10	2-methyl-4,6-dinitrophenol	534-52-1	2.20	198.134	0.27	Fent and Hunn, 1996
11	2-nitrophenol	88-75-5	1.68	139.110	-0.08	Fent and Hunn, 1996
12	2-*s*-butyl-4,6-dinitrophenol	88-85-7	3.36	240.215	0.77	Fent and Hunn, 1996
13	2-*t*-butyl-4,6-dinitrophenol	1420-07-1	3.42	240.215	0.82	Fent and Hunn, 1996
14	3,4-dichlorophenol	95-77-2	2.76	163.003	0.12	Fent and Hunn, 1996
15	3-amino-4-hydroxybenzene sulfonic acid	98-37-3	0.00	189.191	-0.98	Fent and Hunn, 1996
16	3-nitrobenzene sulfonic acid	98-47-5	1.23	203.174	-2.54	Fent and Hunn, 1996
17	3-nitrophenol	554-84-7	1.68	139.110	-0.63	Fent and Hunn, 1996
18	4-aminonaphthalene- 1 - sulfonic acid	84-86-6	1.39	223.252	-2.69	Fent and Hunn, 1996
19	4-chloro-2-methylphenol	1570-64-5	2.69	142.585	0.00	Fent and Hunn, 1996
20	4-chlorophenol	106-48-9	2.20	128.558	-0.49	Fent and Hunn, 1996
21	4-nitrophenol	100-02-7	1.68	139.110	-0.40	Fent and Hunn, 1996
22	4-nonylphenol	104-40-5	5.47	220.356	0.16	Fent and Hunn, 1996
23	4-octylphenol	1806-26-4	5.05	206.329	0.21	Fent and Hunn, 1996
24	Atorvastatin	134523-00-5	3.41	558.650	1.11	Caminada et al., 2006
25	Bezafibrate	41859-67-0	3.77	361.825	-0.24	Caminada et al., 2006
26	Clofibric acid	882-09-7	2.62	214.648	-0.50	Caminada et al., 2007
27	Cyproconazole	94361-06-5	2.79	293.798	0.20	Knauer et al., 2007
28	Cyprodinil	121552-61-2	3.99	225.295	0.59	Knauer et al., 2007
29	Diafenthiuron	80060-09-9	6.36	384.588	2.22	Knauer et al., 2007
30	Diazepam	439-14-5	3.27	298.773	0.36	Caminada et al., 2006
31	Dichlornitroacetanilide	38411-17-5	2.09	249.053	0.46	Knauer et al., 2007

Table 4.8. Continued.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,NR[PLHC-1]}$ (mM)	References
33	Difenoconazole	119446-68-3	4.36	406.269	1.21	Knauer et al., 2007
34	Doxorubicin	23214-92-8	-0.29	527.526	2.66	Caminada et al., 2006
35	Fenpiclonile	74738-17-3	2.64	237.089	0.89	Knauer et al., 2007
36	Fluoxetine	54910-89-3	4.44	309.331	1.62	Caminada et al., 2006
37	Furosemide	54-31-9	-0.14	330.748	-0.61	Caminada et al., 2006
38	Gemfibrozil	25812-30-0	4.41	250.338	0.07	Caminada et al., 2006
39	Hydrochlorothiazide	58-93-5	-3.89	297.743	-0.38	Caminada et al., 2006
40	Ibuprofen	15687-27-1	3.75	206.285	0.08	Caminada et al., 2006
41	Imazalil	35554-44-0	3.57	297.185	0.89	Knauer et al., 2007
42	Mefenamic acid	61-68-7	3.52	241.290	0.66	Caminada et al., 2006
43	Myclobutanil	88671-89-0	4.15	288.782	0.32	Knauer et al., 2007
44	Naphthalene-1,5-disulfonic acid	81-04-9	1.35	288.300	-2.65	Fent and Hunn, 1996
45	Naproxen	22204-53-1	2.97	230.263	-0.17	Caminada et al., 2006
46	OH-Tamoxifen	68047-06-3	5.67	387.523	2.33	Caminada et al., 2006
47	Penconazole	66246-88-6	4.36	286.206	1.00	Knauer et al., 2007
48	Pentachlorophenol	87-86-5	4.44	266.338	0.70	Fent and Hunn, 1996
49	Phenol	108-95-2	1.64	94.113	-0.86	Fent and Hunn, 1996
50	Pravastatin	81093-37-0	2.04	424.534	-0.78	Caminada et al., 2006
51	Prochloraz	67747-09-5	3.78	376.671	0.96	Knauer et al., 2007
52	Propiconazole	60207-90-1	4.24	342.226	0.74	Knauer et al., 2007
53	Propranolol	525-66-6	0.36	259.349	0.85	Caminada et al., 2006
54	Salicylic acid	69-72-7	1.20	138.122	-0.94	Caminada et al., 2006
55	Simvastatin	79902-63-9	4.56	418.574	0.76	Caminada et al., 2006
56	Tamoxifen	10540-29-1	6.05	371.524	1.71	Caminada et al., 2006
57	Tebuconazole	107534-96-3	4.09	307.825	0.59	Knauer et al., 2007

*_n_-octanol-water partition coefficients calculated by SPARTAN 10; ** MW calculated by SPARTAN 10.

4.2.2. Model Development

After numerous training/test set divisions, three descriptor models were generated for the $pEC_{50,NR[PLHC-1]}$ endpoint. Models in terms of fit, internal and external validation parameters were selected from one of the random divisions. Test set chemicals in this random division are given in Table 4.9.

Table 4.9. Test set compounds used in the QSTR modelling of $pEC_{50,NR[PLHC-1]}$.

n_{TEST}/n_{TR}	Test Set Compounds*
18/39	1, 4, 6, 8, 10, 11, 13, 15, 19, 23, 28, 29, 36, 37, 41, 45, 53, 57

*Chemical numbers refer to the ID numbers given in Table 4.8.

Fit and internal validation parameters of the generated one to three-descriptor models for $pEC_{50,NR[PLHC-1]}$ are given in Table 4.10, whereas external validation parameters are given inTable 4.11. The increase in R^2 with the addition of a new variable can be seen in Tables 4.10 and 4.11. Figure 4.6 shows how Q^2_{LOO} and R^2 of the models change with the addition of a new variable. Considering this and other statistical criteria, only three descriptor models for each division were used for further analysis.

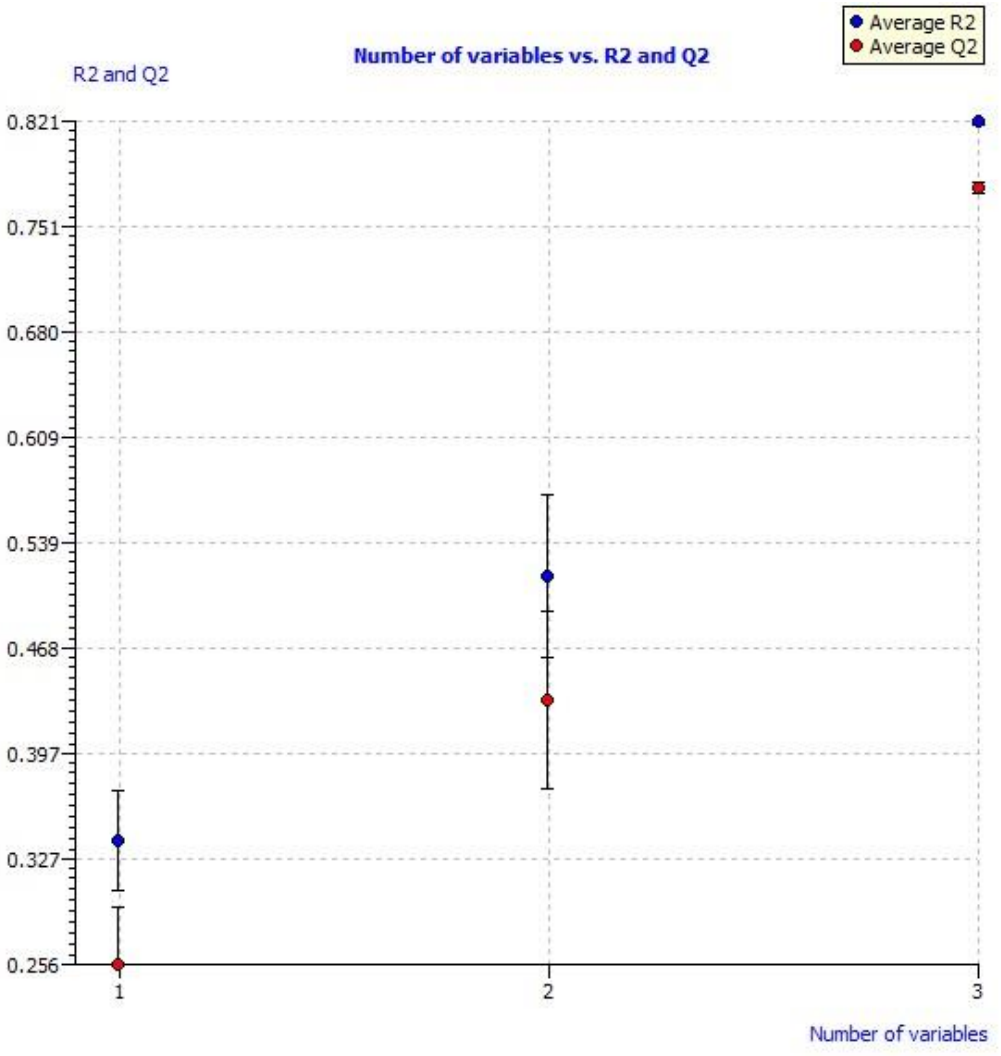


Figure 4.6. The number of variables and the change in R^2 and Q^2_{LOO} for the generated $pEC_{50,NR[PLHC-1]}$ models.

Table 4.10. Fit and internal validation parameters of the generated $pEC_{50,NR[PLHC-1]}$ models.

No	Size	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
***1**	**3**	**CATS2D_01_DN NaasC DLS_04**	**0.825**	**0.810**	**0.466**	**0.904**	**55.148**	**0.782**	**0.520**	**0.882**
2	3	CATS2D_01_DN Mor28u NaasC	0.824	0.809	0.468	0.904	54.650	0.781	0.522	0.881
3	3	CATS2D_01_DN Mor28e NaasC	0.820	0.804	0.474	0.901	52.958	0.772	0.532	0.876
4	3	CATS2D_01_DN Mor28i NaasC	0.816	0.801	0.478	0.899	51.870	0.771	0.534	0.875
5	2	CATS2D_01_DN NaasC	0.612	0.591	0.695	0.760	28.431	0.536	0.760	0.710
6	2	CATS2D_01_DN Mor28u	0.493	0.464	0.795	0.660	17.469	0.406	0.860	0.599
7	2	CATS2D_01_DN Mor28i	0.492	0.464	0.795	0.660	17.443	0.407	0.859	0.599
8	2	CATS2D_01_DN Mor28e	0.471	0.442	0.812	0.640	16.023	0.385	0.875	0.578
9	1	NaasC	0.374	0.357	0.883	0.545	22.114	0.295	0.937	0.494
10	1	CATS2D_01_DN	0.306	0.287	0.930	0.468	16.284	0.217	0.988	0.391

*The selected model.

Table 4.11. External validation parameters of the generated $pEC_{50,NR[PLHC-1]}$ models.

No	Descriptors	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2-r_0'^2)/r^2$	$(r^2-r_0^2)/r^2$
***1**	**CATS2D_01_DN NaasC DLS_04**	**0.799**	**0.360**	**0.800**	**0.782**	**0.896**	**0.892**	**0.757**	**0.074**	**0.898**	**0.926**	**0.000**	**0.012**
2	CATS2D_01_DN Mor28u NaasC	0.761	0.409	0.742	0.720	0.866	0.861	0.655	0.082	0.827	0.942	0.010	0.049
3	CATS2D_01_DN Mor28e NaasC	0.761	0.409	0.742	0.719	0.866	0.864	0.662	0.082	0.856	0.915	0.008	0.044
4	CATS2D_01_DN Mor28i NaasC	0.767	0.405	0.747	0.725	0.869	0.864	0.656	0.080	0.828	0.946	0.011	0.051
5	CATS2D_01_DN NaasC	0.258	0.697	0.250	0.184	0.610	0.459	0.204	0.027	0.394	0.898	0.095	0.264
6	CATS2D_01_DN Mor28u	0.719	0.417	0.731	0.708	0.860	0.847	0.669	0.041	0.851	0.910	0.014	0.003
7	CATS2D_01_DN Mor28i	0.712	0.426	0.720	0.696	0.855	0.844	0.662	0.020	0.859	0.894	0.010	0.005
8	CATS2D_01_DN Mor28e	0.710	0.426	0.720	0.696	0.855	0.841	0.657	0.063	0.856	0.896	0.020	0.001
9	NaasC	0.013	0.914	-0.290	-0.403	0.329	0.086	0.010	0.006	0.074	0.277	0.078	18.480
10	CATS2D_01_DN	0.319	0.637	0.373	0.318	0.674	0.486	0.239	0.139	0.442	1.031	0.693	0.004

*The selected model.

All the selected models demonstrated a good statistical performance: high Q^2_{LOO} values indicate stability and robustness of the models, low $RMSE_{TR}$ and $RMSE_{TEST}$ values show good internal and external performances, respectively, and high CCC_{TEST} and Q^2_{Fn} values propose a high external predictive ability.

4.2.3. Comparison of the Applicability Domain of $pEC_{50,NR[PLHC-1]}$ Models

In an attempt to select the best QSTR model for the $pEC_{50,NR[PLHC-1]}$ endpoint, applicability domain (AD) of four 3-descriptor models were tested to predict $pEC_{50,NR[PLHC-1]}$ values of an external set of 660 chemicals from various classes. Structural coverage of each model is given in Table 4.12.

Table 4.12. Predictive performances of the generated $pEC_{50,NR[PLHC-1]}$ models.

Model no	Number of chemicals within the AD of models (out of 660)	Structural coverage (%)
***1**	**651**	**98.6**
2	646	97.8
3	647	98.0
4	648	98.2

*The selected model.

Model 1 (Eq. 4.3) was selected as the best QSTR model for $pEC_{50,NR[PLHC-1]}$ regarding the structural coverage, statistical parameters addressed and MCDM scores. The numbers in parenthesis indicate the 95% confidence interval of the coefficients. Fit, internal and external validation parameters and structural coverage of Eq. 4.3 are written in bold through Tables 4.10 - 4.12.

$$pEC_{50,NR[PLHC-1]} = 0.393\ (\pm0.092)\ NaasC - 1.182\ (\pm0.311)\ CATS2D\_01\_DN - 2.080\ (\pm0.646)\ DLS\_04 - 2.047\ (\pm0.567) \quad (4.3)$$

The selected model fulfilled all the fit, internal and external validation criteria and it was superior to other models in terms of both internal and external validation parameters. High $R^2 = 0.825$, $R^2_{adj} = 0.810$ and low $RMSE_{TR} = 0.466$ values guaranteed a good fit. Robustness and stability of the model were verified with the satisfying value of Q^2_{LOO} (0.782). As R^2_{Yscr} and Q^2_{Yscr} values were very low (0.079 and -0.150, respectively) the possibility of any chance correlation was eliminated. The predictive performance of the model was identified by high $R^2_{TEST} = 0.799$ and low $RMSE_{TEST} = 0.360$

values. Additionally, the predictive ability of the model was further supported by the external validation parameters Q^2_{F1}, $Q^2_{F2,}$ Q^2_{F3} and CCC_{TEST} for which the model had higher values than the corresponding literature threshold values. The model demonstrated satisfying performances for all the other external validation parameters specified (Table 4.11). Moreover, when Golbraikh and Tropsha's criteria were applied, the model had also a satisfying performance regarding all of the expected conditions. The predictive success of the model was supported by MAE-based criteria (Roy et al., 2016) and the model classified as "good" with values of *MAE* (95% of the data) = 0.306; 3σ = 0.458 and training set range = 5.350.

Chemicals used for the QSTR modelling of $pEC_{50,NR[PLHC-1]}$ endpoint together with their training/test set status in modelling, experimental and predicted $pEC_{50,NR[PLHC-1]}$ values, hat and descriptor values are given in Table 4.13.

Table 4.13. Chemicals used for the QSTR modelling of $pEC_{50,NR[PLHC-1]}$ their experimental and predicted $pEC_{50,NR[PLHC-1]}$ values, hat and descriptor values.

ID	Name	Status	Exp. $pEC_{50,NR[PLHC-1]}$ (mM)	NaasC	CATS2D_01_DN	DLS_04	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq. 4.3 (mM)	Hat value (h*=0.308)
1	*(S)*-propranolol	Prediction	0.74	1	0	1	0.426	0.136
2	2,3,4,5-tetrachlorophenol	Training	0.77	5	0	0.3	0.541	0.087
3	2,3,6-trichlorophenol	Training	-0.18	4	0	0.3	0.148	0.074
4	2,4,5-trichlorophenol	Prediction	0.66	4	0	0.3	0.148	0.074
5	2,4-dichlorophenol	Training	0.55	3	0	0.3	-0.244	0.078
6	2,4-dinitrophenol	Prediction	0.09	3	0	0.4	-0.036	0.056
7	2,5-dinotrophenol	Training	0.28	3	0	0.4	-0.036	0.056
8	2,6-dichlorophenol	Prediction	-0.72	3	0	0.3	-0.244	0.078
9	2-chlorophenol	Training	0.08	2	0	0.5	-0.221	0.060
10	2-methyl-4,6-dinitrophenol	Prediction	0.27	4	0	0.4	0.356	0.054
11	2-nitrophenol	Prediction	-0.08	2	0	0.4	-0.429	0.075
12	2-*s*-butyl-4,6-dinitrophenol	Training	0.77	4	0	0.8	1.188	0.057
13	2-*t*-butyl-4,6-dinitrophenol	Prediction	0.82	4	0	0.8	1.188	0.057
14	3,4-dichlorophenol	Training	0.12	3	0	0.3	-0.244	0.078
15	3-amino-4-hydroxybenzene sulfonic acid	Prediction	-0.98	3	1	0.4	-1.218	0.099
16	3-nitrobenzene sulfonic acid	Training	-2.54	2	1	0.4	-1.611	0.113
17	3-nitrophenol	Training	-0.63	2	0	0.4	-0.429	0.075
18	4-aminonaphthalene- 1 -sulfonic acid	Training	-2.69	2	1	0.4	-1.611	0.113
19	4-chloro-2-methylphenol	Prediction	0	3	0	0.5	0.172	0.043
20	4-chlorophenol	Training	-0.49	2	0	0.5	-0.221	0.060
21	4-nitrophenol	Training	-0.4	2	0	0.4	-0.429	0.075
22	4-nonylphenol	Training	0.16	2	0	0.8	0.403	0.065
23	4-octylphenol	Prediction	0.21	2	0	0.9	0.611	0.084

Table 4.13. Continued.

ID	Name	Status	Exp. $pEC_{50,NR[PLHC-1]}$ (mM)	NaasC	CATS2D_01_DN	DLS_04	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq. 4.3 (mM)	Hat value (h*=0.308)
24	Atorvastatin	Training	1.11	8	1	0.7	1.370	0.276
25	Bezafibrate	Training	-0.24	4	1	1	0.423	0.125
26	Clofibric acid	Training	-0.5	2	1	1	-0.363	0.116
27	Cyproconazole	Training	0.2	2	0	1	0.819	0.110
28	Cyprodinile	Prediction	0.59	4	0	0.8	1.188	0.057
29	Diafenthiuron	Prediction	2.22	5	0	1	1.997	0.134
30	Diazepam	Training	0.36	4	0	0.6	0.772	0.039
31	Dichlornitroacetanilide	Training	0.46	4	0	0.4	0.356	0.054
32	Diclofenac	Training	0.57	5	1	0.6	-0.016	0.099
33	Difenoconazole	Training	1.21	5	0	0.8	1.581	0.079
34	Doxorubicin	Training	2.66	8	0	0.7	2.551	0.224
35	Fenpiclonile	Training	0.89	5	0	0.4	0.749	0.069
36	Fluoxetine	Prediction	1.62	3	0	1	1.212	0.101
37	Furosemide	Prediction	-0.61	5	1	0.4	-0.432	0.122
38	Gemfibrozil	Training	0.07	3	1	1	0.030	0.112
39	Hydrochlorothiazide	Training	-0.38	4	0	0.3	0.148	0.074
40	Ibuprofen	Training	0.08	2	1	1	-0.363	0.116
41	Imazalile	Prediction	0.89	3	0	0.8	0.796	0.053
42	Mefenamic acid	Training	0.66	5	1	0.6	-0.016	0.099
43	Myclobutanil	Training	0.32	2	0	0.8	0.403	0.065
44	Naphthalene-1,5-disulfonic acid	Training	-2.65	2	2	0.4	-2.792	0.344
45	Naproxen	Prediction	-0.17	2	1	0.8	-0.779	0.081
46	OH-tamoxifen	Training	2.33	5	0	0.8	1.581	0.079
47	Penconazole	Training	1	3	0	1	1.212	0.101
48	Pentachlorophenol	Training	0.7	6	0	0.3	0.934	0.117

Table 4.13. Continued.

ID	Name	Status	Exp. $pEC_{50,NR[PLHC-1]}$ (mM)	NaasC	CATS2D_01_DN	DLS_04	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq. 4.3 (mM)	Hat value (h*=0.308)
49	Phenol	Training	-0.86	1	0	0.5	-0.614	0.095
50	Pravastatin	Training	-0.78	0	1	1	-1.148	0.174
51	Prochloraz	Training	0.96	4	0	0.8	1.188	0.057
52	Propiconazole	Training	0.74	3	0	0.8	0.796	0.053
53	Propranolol	Prediction	0.85	1	0	1	0.426	0.136
54	Salicylic acid	Training	-0.94	2	1	0.4	-1.611	0.113
55	Simvastatin	Training	0.76	0	0	1	0.034	0.179
56	Tamoxifen	Training	1.71	4	0	0.7	0.980	0.044
57	Tebuconazole	Prediction	0.59	2	0	1	0.819	0.110

The linear relationship between predicted $pEC_{50,NR[PLHC-1]}$ values from Eq 4.3 versus experimental values $pEC_{50,NR[PLHC-1]}$ is given in Figure 4.7. A good alignment of the data along the optimal line can be observed hence it can be concluded that there is an agreement between the predicted $pEC_{50,NR[PLHC-1]}$ with the experimental $pEC_{50,NR[PLHC-1]}$ values.

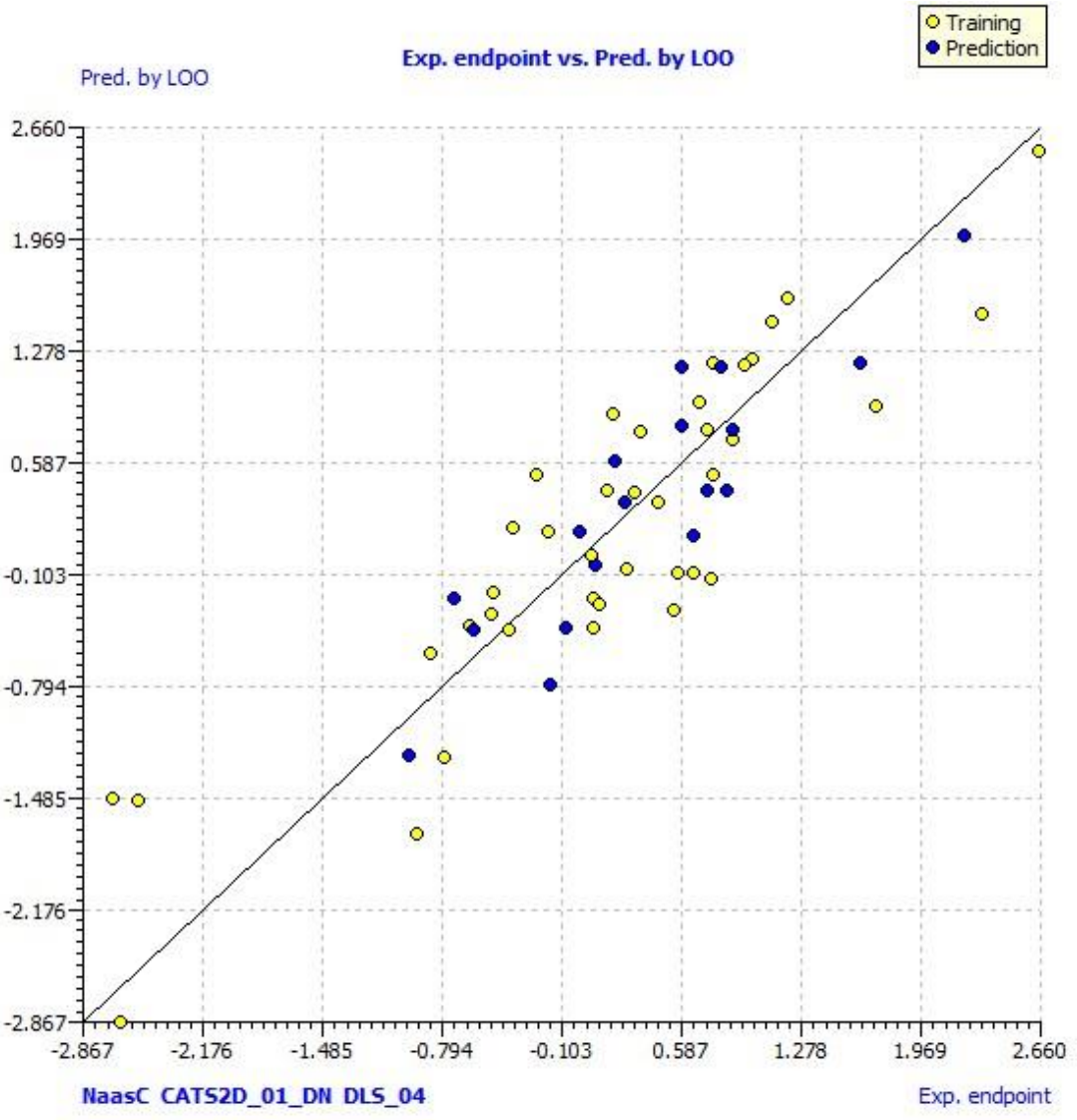


Figure 4.7. Plot of predicted $pEC_{50,NR[PLHC-1]}$ values from Eq. 4.3 versus experimental $pEC_{50,NR[PLHC-1]}$ values.

Williams plot (standardized residuals versus hat values), indicating the AD of the $pEC_{50,NR[PLHC-1]}$ model, can be visualized in Figure 4.8. It is important to state that all the training and test set compounds had residuals lower than ±2.5σ hence no response outliers were present. However, naphthalene-1,5-disulfonic acid was a structural outlier with a hat value 0.344 greater than the critical hat value (h^* = 0.308). Naphthalene-1,5-disulfonic acid is a structurally different chemical from the

rest of the data set as it is the only compound including naphthalene skeleton. Moreover, considering the descriptors in Eq. 4.3, naphthalene-1,5-disulfonic acid has the highest drug-like score, DLS_04 among the data set chemicals thus its structural difference can be further emphasized.

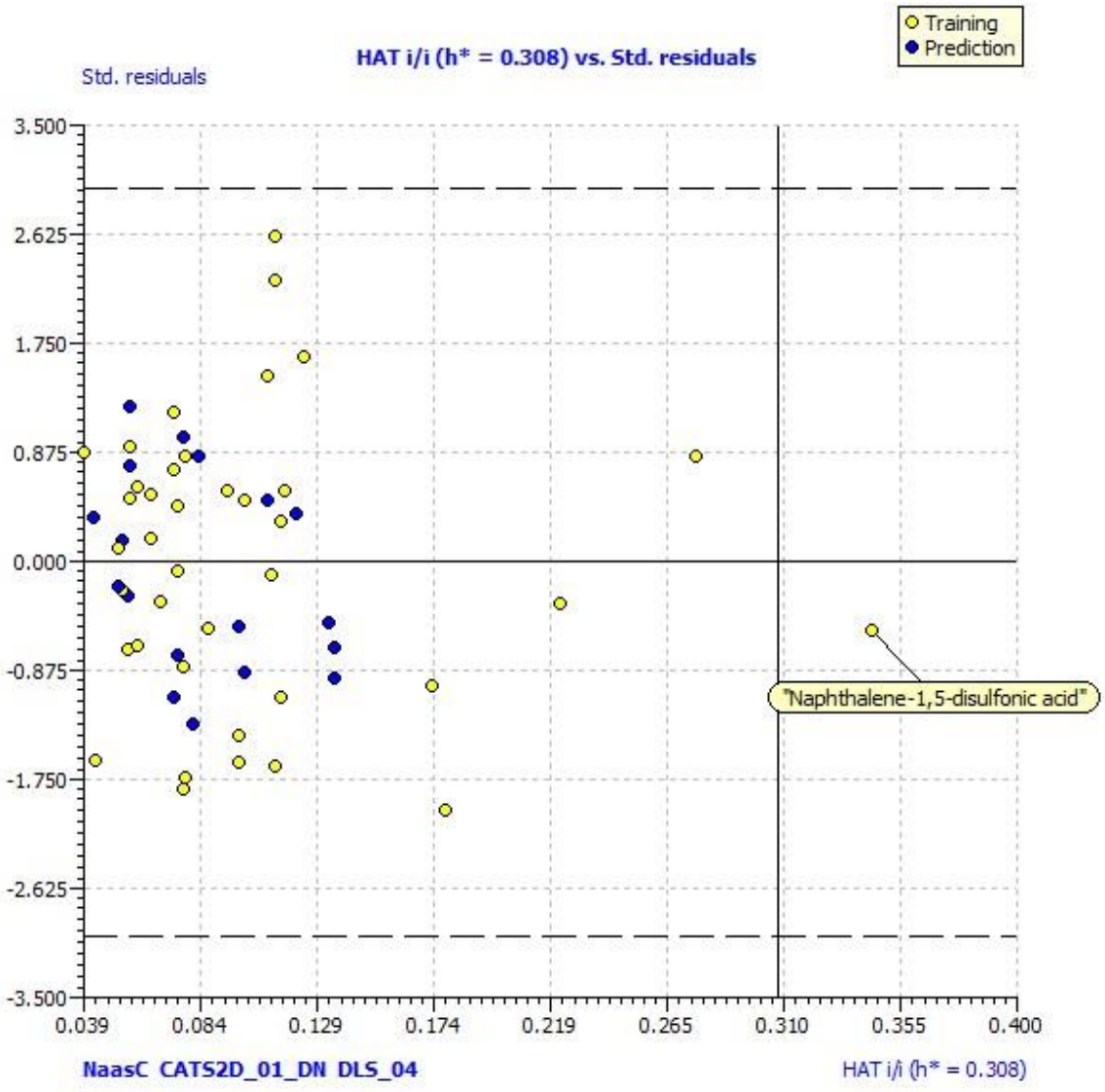


Figure 4.8. Williams plot of $pEC_{50,NR[PLHC-1]}$ model.

To test the external predictive ability of the generated $pEC_{50,NR[PLHC-1]}$ model, an external set composed of 660 chemicals was used. External set chemicals, calculated descriptor, predicted and experimental $pEC_{50,NR[PLHC-1]}$ values are provided in Appendix B. The model had 98.6% structural coverage which can be seen in the Insubria graph (Figure 4.9) where predicted $pEC_{50,NR[PLHC-1]}$ values for training, test and external set chemicals versus their hat values are depicted.

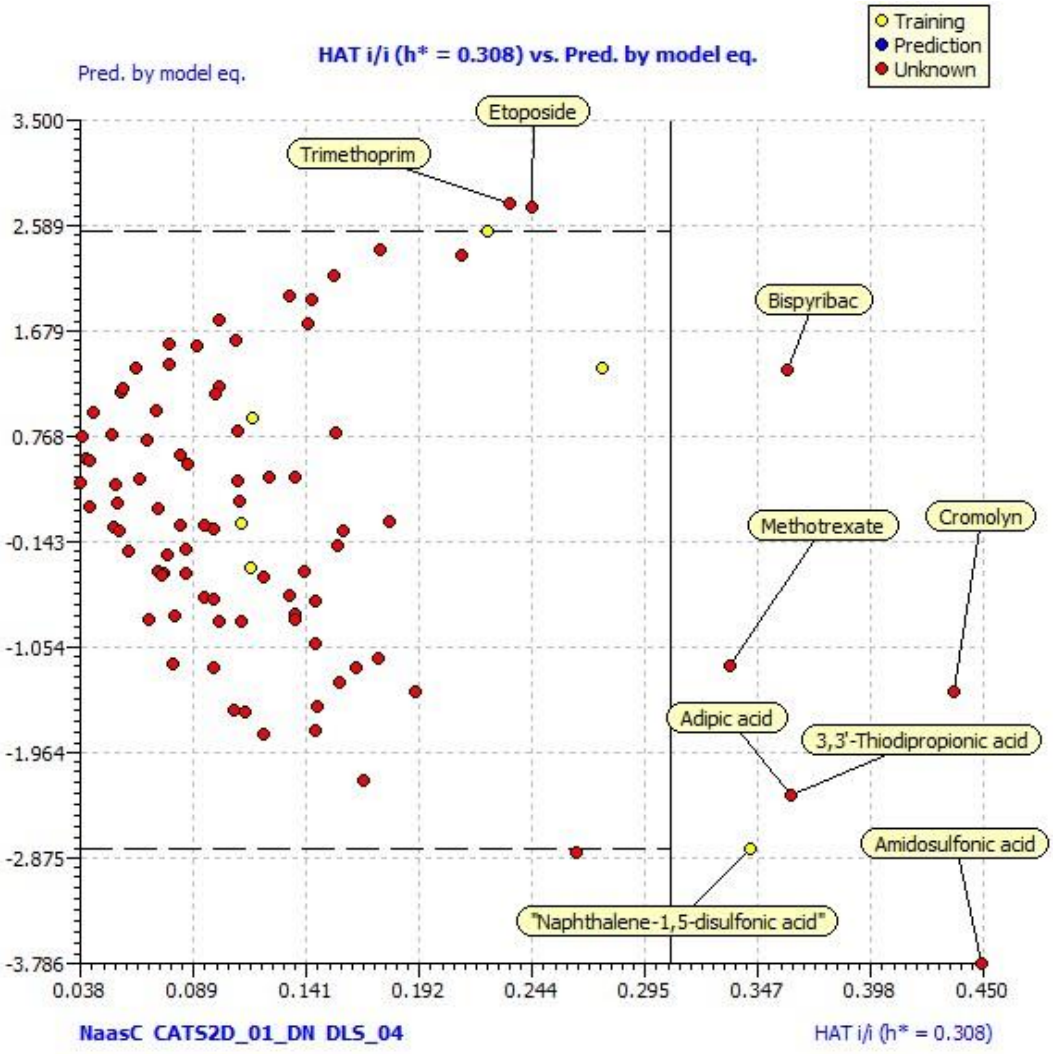


Figure 4.9. Insubria graph of $pEC_{50,NR[PLHC-1]}$ model. Predicted $pEC_{50,NR[PLHC-1]}$ values for training, test and external (660 chemicals) set chemicals from Eq. 4.3 versus their hat values.

Ten chemicals with the most and the least $pEC_{50,NR[PLHC-1]}$ values depending on the predictions from Eq. 4.3 were screened and given in Table 4.13. The chemical with highest $pEC_{50,NR[PLHC-1]}$ value is trichlorosyringol which is a chlorinated phenol. It is disposed into the environment as a residue of bleaching process of pulp in the paper industry (Paasivirta, 1995). Regarding the high cytotoxicity reported in the present study, it would be beneficial to screen its disposal by means of legal regulations. When the classes of the chemicals in Table 4.14 were examined, there were five pharmaceuticals and three pesticides out of ten. Therefore, the presented results might serve as an initial prioritisation for pharmaceuticals and pesticides and may emphasize the significance of assessment and screening for this type of chemicals.

Molecular descriptors appeared in Eq. 4.3 are given in Table 4.15 together with their standardized coefficients. Standardized coefficients show the importance of a descriptor in the model relative to the other descriptors.

The most important descriptor appeared in the model is NaasC, an Electrotopological State (E-state) index descriptor which positively correlated with the modeling endpoint. E-state indices are encoded by an intrinsic electronic state of a specific atom in a molecule. Intrinsic state is a value that is obtained from the combination of electronic distribution and topological accessibility. Intrinsic state (I) value is formulated as:

$$I=\frac{[(2/N)^2]\delta^V+1}{\delta} \tag{4.4}$$

In this formula, δ^V represents the total number of valence electrons, δ represents the number of skeletal neighbors and N is the principal quantum number (Hall, 2004).

To accurately formulate E-state indices, the influence of other atoms with the target atom must be considered besides the intrinsic state of the atom. For this purpose, perturbations from each atom are calculated with the formula given in Eq. 4.5.

$$\Delta I_{ij}=\frac{(I_i-I_j)}{r_{ij}^2} \tag{4.5}$$

where (I_i-I_j) is intrinsic state difference and r_{ij} is the graph distance between the target atom and other atoms.

Table 4.14. The most and the least cytotoxic chemicals from the external set predicted by Eq. 4.3.

	Name	Chemical class	$pEC_{50,\ NR[PLHC-1]}$ pred. by Eq. 4.3 (mM)
Most Toxic	Trichlorosyringol	Phenol, chloro	2.390
	Tetrabromobisphenol A	Diphenyl alkane, dihydroxy, bromo	2.343
	Novobiocin	Pharmaceutical, Anti-bacterial Agents	2.159
	Glibenclamide	Pharmaceutical, Antidiabetic Agents	1.997
	Metipranolol	Pharmaceutical, Beta-adrenergic Antagonist	1.997
	Pendimethalin	Pesticide, Herbicide	1.997
	Midazolam	Pharmaceutical, Benzodiazepines, Antianxiety Agents	1.951
	Ketoconazole	Pharmaceutical, Antifungal Agents	1.789
	Pyrazosulfuron ethyl	Pesticide, Herbicide	1.789
	Cyclosulfamuron	Pesticide, Herbicide	1.789
Least Toxic	Acrylic acid	Acid, a,b-unsaturated	-2.188
	L-tryptophan	Pharmaceutical, Antidepressive Agents	-1.795
	Perfluorooctane sulfonic acid	Sulfonic acid, polyfuoro	-1.772
	3-hydroxy-2-naphthoic acid	Naphthalene, acid-hydroxy	-1.611
	4-hydroxybenzoic acid	Benzoic acid, hydroxy	-1.611
	Cefuroxime	Pharmaceutical, Antiinfective Agents	-1.587
	Trans-cinnamic acid	Benzene-acid, a,b-unsaturated	-1.587
	Decanoic acid	Pesticide, Insecticide, Acaricide, Herbicide, Plant Growth Regulator	-1.564
	Octanoic acid	Acid	-1.564
	Perfluorooctanoic acid	Acid, polyfluoro	-1.564

Therefore, E-state of each atom is given as the combination of intrinsic state value and perturbations from all other atoms (Hall, 2004) (Eq. 4.6).

$$S_i = I_i + \sum_i I_{ij} \quad (4.6)$$

Table 4.15. Descriptors appeared in the $pEC_{50,NR[PLHC-1]}$ model.

Descriptor	Meaning of descriptor	Type	Standardized coefficient
NaasC	Number of atoms of type aasC Atom-type	E-state indices	0.623
CATS2D_01_DN	CATS2D Donor-Negative at lag 01	CATS 2D	-0.554
DLS_04	Modified drug-like score from Chen et al. (7 rules)	Drug-like indices	0.472

In an attempt to explain the presence of NaasC in the $pEC_{50,NR[PLHC-1]}$ model, it would be useful to state that the endpoint measures the uptake of NR inside a target cell. As the descriptor is related to electrochemical properties of a molecule, it is likely that the electrochemical processes occur during the exposure of chemical in lysosomes.

E-state indices have often be used in QSAR modeling studies (Hall and Kier, 1995) for example, structure-activity relationships of toxicity of diverse set of chemicals to *Pimephales promelas* by Gramatica and co-workers (2014); microbial mutagenicity and carcinogenic potential of pharmaceuticals in rodents by Contrera and co-workers (2003).

The second descriptor was CATS2D_01_DN that belongs to the CATS 2D descriptor block. CATS 2D descriptors are holographic vector representations of molecules encoding topological distances between any pair of pharmacophore points (Todeschini and Consonni, 2009). There are five defined pharmacophore points: hydrogen-bond donor (D), hydrogen-bond acceptor (A), positively charged or ionizable (P), negatively charged or ionizable (N) and lipophilic (L) and those pharmacophore points combined into 15 pharmacophore pairs (DD, DA, DP, DN, DL, AA, AP, AN, AL, PP, PN, PL, NN, NL, LL).

For descriptor calculation, pharmacophore pairs and topological distances between all pairs in a molecule are calculated. Then, each pharmacophore from 15 combinations is assigned to a number of bins which encode the number of occurrence of a potential pharmacophore pair (Todeschini and

Consonni, 2009). Along with occurrence of 15 pharmacophore pairs and the topological distance between intervening bonds which are assigned with numbers 0 to 9 resulting in a 150-dimensional autocorrelation vector.

In CATS2D_01_DN, DN refers to the hydrogen-bond donor (D), either negatively charged or ionizable (N) (Todeschini and Consonni, 2009). It was inversely related to the $pEC_{50,NR[PLHC-1]}$ endpoint values. The presence of this descriptor in the model can be attributed to the possible effect of specific position and charge of hydrogen-bond donors in a molecule on the potency of that molecule to cause lysosomal damage.

The last descriptor, DLS_04 is a drug-like descriptor. It was the least significant descriptor in the model and contributed negatively to the lysosomal damage ($pEC_{50,NR[PLHC-1]}$) endpoint. The descriptor as modified by Chen and co-workers (2005) is calculated using seven physicochemical properties: $logK_{ow}$, molecular weight, hydrogen-bond acceptor, hydrogen-bond donor, ratio of the number of C (sp^3) atoms to the total number of non-halogen heavy atoms, ratio of the number of hydrogen atoms to the total number of non-halogen heavy atoms, ratio of the molecular unsaturation to the total number of non-halogen heavy atoms. As the descriptor encodes information about various physicochemical properties, they should be considered in explaining lysosomal damage.

In a nanoSAR study by Oksel and co-workers (2016), DLS_04 was found to be effective in modelling of cellular uptake of 105 iron-oxide based nanoparticles in pancreatic cancer cells.

4.2.4. Correlation between $pEC_{50,NR[PLHC-1]}$ and log K_{ow}

The log K_{ow} values ranged from -3.89 for hydrochlorothiazide to 6.36 for diafenthiuron. The regression analysis revealed that there is a moderate correlation between log K_{ow} and $pEC_{50,NR[PLHC-1]}$ values for all compounds ($n = 57$; $r = 0.464$; $p < 0.01$). The plot of observed $pEC_{50,NR[PLHC-1]}$ values versus log K_{ow} values is given in Figure 4.10. In the study by Fent and Hunn (1996) a good correlation was reported between log K_{ow} and $pEC_{50,NR[PLHC-1]}$ values for 25 chemicals. However, it can be claimed that with the addition of structurally divergent chemicals (i.e., pesticides and pharmaceuticals), the data set became more complex to explain $pEC_{50,NR[PLHC-1]}$ endpoint using a single variable.

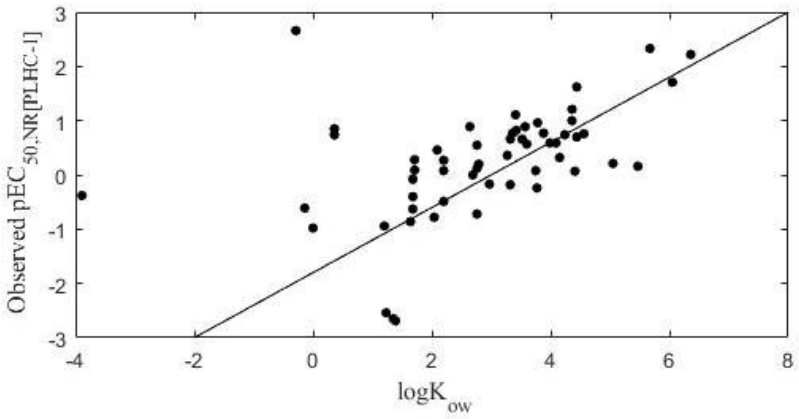


Figure 4.10. Plot of observed $pEC_{50,NR[PLHC-1]}$ values versus log K_{ow} values.

4.2.5. Correlation of Experimental *in vivo* pLC_{50} with Predicted *in vitro* $pEC_{50,NR[PLHC-1]}$

To assess the suitability of the proposed model as an alternative of *in vivo* toxicity testing, the relationship between the predicted $pEC_{50,NR[PLHC-1]}$ values and the fish lethality (pLC_{50}) data compiled from the literature was evaluated. The experimental pLC_{50} and predicted $pEC_{50,NR[PLHC-1]}$ values for external set chemicals are given in Appendix B.

Correlation analysis revealed a moderate correlation between the predicted *in vitro* ($pEC_{50,NR[PLHC-1]}$ values) and *in vivo* (pLC_{50}) data (n = 179; r = 0.502; p < 0.01). Plot of observed pLC_{50} versus predicted $pEC_{50,NR[PLHC-1]}$ values is given in Figure 4.11.

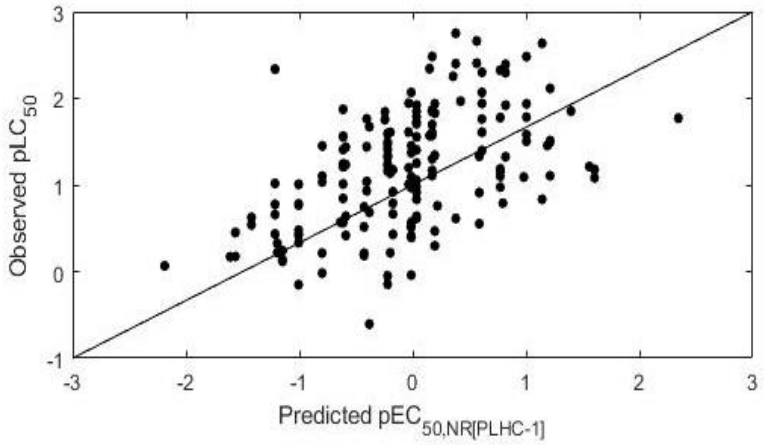


Figure 4.11. Plot of observed pLC_{50} versus predicted $pEC_{50,NR[PLHC-1]}$ values by Eq. 4.3.

From the 207 external set chemicals having experimental pLC_{50} values, the relationship between the predicted $pEC_{50,NR[PLHC-1]}$ and the experimental pLC_{50} values of 28 chemicals was weak. These

chemicals were particularly from bromo- and chloro- substituted alkanes, and different chemical classes. These chemicals were not included in the correlation analysis.

4.3. QSTR Modelling for $pEC_{50,NR[GFS]}$

4.3.1. Dataset

Dataset chemicals used in $pEC_{50,NR[GFS]}$ model with the experimental neutral red uptake values $pEC_{50,NR[GFS]}$ and the relevant properties are listed in Table 4.16.

Table 4.16. The name, CAS number, log K_{ow}, molecular weight and experimental $pEC_{50,NR[GFS]}$ values of chemicals in the dataset.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,NR[GFS]}$ (mM)
1	1,2,4-trichlorobenzene	120-82-1	3.71	181.449	0.69
2	1,4-dioxane	123-91-1	-0.32***	84.074	-2.73
3	1-butanol	71-36-3	0.97	74.123	-1.81
4	1-naphthol	90-15-3	2.64	144.173	0.47
5	1-octanol	111-87-5	2.64	130.231	-0.26
6	1-propanol	71-23-8	0.55	60.096	-2.18
7	2,3,4,6-tetrachlorophenol	58-90-2	3.88	231.893	1.6
8	2,3,4-trichlorophenol	15950-66-0	3.32	197.448	1.68
9	2,3-dichlorophenol	576-24-9	2.76	163.003	0.66
10	2,4,5-trichlorophenol	95-95-4	3.32	197.448	1.79
11	2,4,6-trichlorophenol	88-06-2	3.32	197.448	0.38
12	2,4,6-triiodophenol	609-23-4	5.72	471.801	1.82
13	2,4-dichlorophenol	120-83-2	2.76	163.003	0.84
14	2,4-dimethylphenol	105-67-9	2.62	122.167	0.04
15	2,5-dichlorophenol	583-78-8	2.76	163.003	0.78
16	2,6-dichlorophenol	87-65-0	2.76	163.003	-0.21
17	2-bromophenol	95-56-7	2.47	173.009	0.23
18	2-butanol	78-92-2	0.87	74.123	-2.13
19	2-butoxyethanol	111-76-2	0.81	118.176	-1.86
20	2-chlorophenol	95-57-8	2.2	128.558	-0.29
21	2-ethoxyethanol	110-80-5	-0.09	90.122	-2.54
22	2-methoxyethanol	109-86-4	-0.43	76.095	-2.9
23	2-methylphenol	95-48-7	2.13	108.14	-0.36
24	2-napthol	135-19-3	2.64	144.173	0.47
25	2-nitrophenol	88-75-5	1.68	139.11	-0.64
26	2-pentanol	6032-29-7	1.29	88.15	-1.78
27	2-propanol	67-63-0	0.38	60.096	-2.71
28	3,4-dichlorophenol	95-77-2	2.76	163.003	1.20

Table 4.16. Continued.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,NR[GFS]}$ (mM)
29	3,5-dichlorophenol	591-35-5	2.76	163.003	1.39
30	3,5-dimethoxyphenol	500-99-2	1.39	154.165	-0.46
31	3,5-dimethylphenol	108-68-9	2.62	122.167	-0.27
32	3,5-di-*tert*-butylphenol	1138-52-9	5.05	206.329	1.68
33	3-bromophenol	591-20-8	2.47	173.009	0.58
34	3-chlorophenol	108-43-0	2.2	128.558	0.07
35	3-methylphenol	108-39-4	2.13	108.14	-0.41
36	4-bromophenol	106-41-2	2.47	173.009	0.31
37	4-chloro-3,5-dimethyl phenol	88-04-0	3.18	156.612	1.03
38	4-chlorophenol	106-48-9	2.20	128.558	-0.11
39	4-methyl-2-Nitrophenol	119-33-5	2.17	153.137	0.01
40	4-methylphenol	106-44-5	2.13	108.14	-0.07
41	4-nitrophenol	100-02-7	1.68	139.11	0.53
42	Acetone	67-64-1	0.20	58.08	-2.42
43	Acetonitrile	75-05-8	0.67	41.053	-3.02
44	Anilazine	101-05-3	3.64***	275.526	1.19
45	Aniline	62-53-3	1.23	93.129	-1.05
46	Benzene	71-43-2	2.03	78.114	-1.24
47	Benzoximate	29104-30-1	4.60	363.797	1.23
48	Bibenzyl	103-29-7	4.54	182.266	1.22
49	Captafol	2425-06-1	3.42***	351.081	1.71
50	Captan	133-06-2	2.74***	296.561	1.66
51	Chlornitrofen	1836-77-7	4.77	318.543	1.95
52	Chlorobenzilate	510-15-6	4.38	325.191	1.76
53	Chlorothalonil	1897-45-6	4.33	265.914	2.26
54	Chlorpyrifos	2921-88-2	5.12***	350.59	1.77
55	Chlorpyrifos-methyl	5598-13-0	4.13***	322.536	1.42
56	Cloroform	67-66-3	1.67	119.378	-1.07
57	Cyanophos	2636-26-2	2.76***	243.223	0.73
58	Diazinon	333-41-5	3.86***	304.351	0.97
59	Dicofol	115-32-2	5.89	370.49	2.18
60	Diethyl ether	60-29-7	0.76	74.123	-2.43
61	di-*n*-butylphthalate	84-74-2	4.16	278.348	1.51
62	Dioxabenzofos	3811-49-2	2.88***	216.197	0.43
63	Diphenyl	92-52-4	3.71	154.212	0.91
64	Disulfoton	298-04-4	4.07***	274.41	0.85
65	Edifenphos	17109-49-8	3.61***	310.378	1.20
66	EPN	2104-64-5	4.47***	323.309	1.82
67	Ethanol	64-17-5	0.07	46.069	-2.35
68	Fenitrothion	122-14-5	3.30***	277.237	1.32
69	Fenobucarb	3766-81-2	3.18	207.273	-0.17
70	Fenthion	55-38-9	4.08***	278.333	1.35
71	gamma-BHC	319-84-6	4.14	290.832	1.21
72	Iprofenfos	26087-47-8	3.57***	288.348	0.38

Table 4.16. Continued.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,NR[GFS]}$ (mM)
73	Isobutanol	78-83-1	0.95	74.123	-1.8
74	Isoprothiolane	50512-35-1	2.58	290.404	0.06
75	Isoxathion	18854-01-8	3.90***	313.314	1.33
76	Malathion	121-75-5	2.29***	330.362	0.43
77	Methanol	67-56-1	-0.27	32.042	-2.7
78	Methidathion	950-37-8	1.58***	302.336	0.27
79	Molinate	2212-67-1	2.15	187.307	-0.2
80	*m*-xylene	108-38-3	3.01	106.168	0.08
81	Naled	300-76-5	1.61***	380.784	1.01
82	Naphthalene	91-20-3	3.03	128.174	0.28
83	Oxadiazon	19666-30-9	5.46	345.226	1.44
84	*o*-xylene	95-47-6	3.01	106.168	0.22
85	p,p-DDD	72-54-8	5.81	320.046	1.79
86	Pentachlorophenol	87-86-5	4.44	266.338	2.17
87	Penthoate	10023-74-2	1.18	102.133	0.89
88	Phenol	108-95-2	1.64	94.113	-0.9
89	Propanil	709-98-8	2.71	218.083	0.33
90	Propaphos	7292-16-2	3.73***	304.347	0.84
91	Propargite	2312-35-8	5.57***	350.479	2.96
92	*p*-xylene	106-42-3	3.01	106.168	-0.23
93	Pyrazolate	58011-68-0	0.58	112.088	1.44
94	Pyridine	110-86-1	0.7	79.102	-1.99
95	Quintozene	82-68-8	4.86	295.336	2.16
96	Salicylaldehyde	90-02-8	1.39	122.123	1.09
97	Simetryn	1014-70-6	1.88	213.309	-0.49
98	*tert*-amyl alcohol	75-85-4	1.09	88.15	-2.05
99	*tert*-butanol	75-65-0	0.6	74.123	-2.38
100	Trifluralin	1582-09-8	4.96	335.282	1.6

**n*-octanol water partition coefficients calculated by SPARTAN 10; **MW calculated by SPARTAN 10; ***calculated by ECOSAR v1.11

4.3.2. Model Development

The number of chemicals in training and test sets together with the composition of test sets used in various divisions for the QSTR modelling of $pEC_{50,NR[GFS]}$ are given in Table 4.17. Best models were selected from four divisions composed of both response-based (division 1) and randomly divided sets (divisions 2, 3, 4).

For each division, 2-4 descriptor models were generated. Fit, internal validation and external validation parameters for these models were given in Tables 4.18 and 4.19, respectively. All the models generated were stable and robust with $R^2 > 0.5$ and $Q^2_{LOO} > 0.5$ and had a good external

performance with $R^2_{TEST} > 0.5$ and $Q^2_{LOO} > 0.7$ for each model (threshold values for Q^2_{Fn} =0.7). Moreover, neither structural nor response outliers were present in the model's ADs.

Table 4.17. The test set compounds for four divisions used in QSTR modelling of $pEC_{50,NR[GFS]}$.

Division no	n_{TEST}/n_{TR}	Test Set Compounds*
1	20/80	1, 4, 15, 16, 17, 18, 22, 23, 45, 48, 60, 61, 68, 70, 73, 85, 86, 88, 89, 98
2	19/81	3, 6, 7, 9, 27, 35, 36, 40, 45, 70, 75, 76, 81, 83, 85, 90, 92, 95, 98
3	26/74	1, 4, 10, 13, 20, 26, 27, 29, 33, 35, 39, 40, 41, 46, 48, 49, 51, 57, 60, 63, 67, 69, 79, 82, 83, 86
4	22/78	6, 9, 15, 17, 18, 34, 36, 37, 39, 40, 42, 44, 45, 51, 52, 55, 59, 66, 67, 70, 75, 81

*Compound numbers refer to the ID numbers given in Table 4.16.

Table 4.18. Fit and internal validation parameters of the generated $pEC_{50,NR[GFS]}$ models.

No	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
				Division 1					
1.1	R2u SP05	0.794	0.789	0.640	0.885	148.381	0.779	0.663	0.877
1.2	R2u DP05	0.787	0.781	0.651	0.881	141.937	0.771	0.675	0.872
1.3	SP05	0.608	0.603	0.883	0.756	120.791	0.589	0.904	0.744
1.4	DP05	0.589	0.584	0.904	0.742	112.021	0.571	0.924	0.729
				Division 2					
***2.1**	**GATS1v DP07**	**0.811**	**0.809**	**0.620**	**0.897**	**169.854**	**0.799**	**0.644**	**0.889**
2.2	DP07	0.561	0.556	0.951	0.719	100.999	0.540	0.974	0.705
2.3	GATS1v	0.441	0.433	1.074	0.612	62.190	0.416	1.097	0.593
				Division 3					
3.1	SpPos_B(v) SpMaxA_G/D	0.800	0.794	0.657	0.889	141.795	0.782	0.685	0.879
3.2	SpPos_B(v)	0.613	0.608	0.914	0.760	114.036	0.591	0.940	0.747
33	SpMaxA_G/D	0.171	0.160	1.337	0.293	14.908	0.119	1.379	0.249
				Division 4					
4.1	X3sol Eig05_AEA(ed) Mor22e O-057	0.820	0.810	0.609	0.901	83.199	0.797	0.646	0.889
4.2	X3sol Eig05_AEA(ed) Mor22u O-057	0.818	0.808	0.612	0.900	82.118	0.795	0.650	0.888
4.3	X3sol Eig05_AEA(ed) Mor22i O-057	0.815	0.805	0.618	0.898	80.246	0.791	0.656	0.885
4.4	Eig05_AEA(ed) O-057	0.760	0.753	0.704	0.863	118.471	0.740	0.733	0.852
4.5	X3sol	0.595	0.590	0.914	0.747	111.875	0.571	0.941	0.732

*The selected model.

Table 4.19. External validation parameters of the generated $pEC_{50,NR[GFS]}$ models.

No	Descriptors	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2–r_0'^2)/r^2$	$(r^2–r_0^2)/r^2$
						Division 1							
1.1	R2u SP05	0.945	0.388	0.937	0.933	0.924	0.963	0.863	0.001	0.876	1.070	0.008	0.008
1.2	R2u DP05	0.937	0.406	0.931	0.927	0.917	0.960	0.864	0.002	0.871	1.069	0.006	0.007
1.3	SP05	0.771	0.771	0.750	0.735	0.701	0.825	0.701	0.003	0.620	1.232	0.010	0.011
1.4	DP05	0.743	0.806	0.727	0.710	0.673	0.807	0.680	0.004	0.601	1.226	0.009	0.010
						Division 2							
***2.1**	**GATS1v DP07**	**0.977**	**0.226**	**0.975**	**0.975**	**0.975**	**0.988**	**0.968**	**0.002**	**1.013**	**0.964**	**0.000**	**0.000**
2.2	DP07	0.669	0.833	0.665	0.664	0.664	0.795	0.601	0.036	0.661	1.007	0.025	0.008
2.3	GATS1v	0.641	0.860	0.643	0.641	0.642	0.769	0.633	0.008	0.600	0.008	0.01	0.000
						Division 3							
3.1	SpPos_B(v) SpMaxA_G/D	0.905	0.425	0.899	0.899	0.916	0.943	0.887	0.012	0.837	1.083	0.000	0.001
3.2	SpPos_B(v)	0.587	0.885	0.561	0.561	0.637	0.710	0.511	0.020	0.525	1.095	0.021	0.037
3.3	SpMaxA_G/D	0.527	1.033	0.402	0.402	0.506	0.486	0.313	0.140	0.313	1.479	0.554	0.143
						Division 4							
4.1	X3sol Eig05_AEA(ed) Mor22e O-057	0.958	0.293	0.958	0.958	0.958	0.978	0.937	0.009	0.960	1.000	0.001	0.000
4.2	X3sol Eig05_AEA(ed) Mor22u O-057	0.965	0.267	0.965	0.965	0.965	0.982	0.945	0.008	0.963	1.004	0.001	0.000
4.3	X3sol Eig05_AEA(ed) Mor22i O-057	0.967	0.262	0.967	0.966	0.967	0.983	0.938	0.008	0.971	0.997	0.001	0.001
4.4	Eig05_AEA(ed) O-057	0.958	0.300	0.956	0.956	0.956	0.977	0.939	0.009	0.928	1.034	0.000	0.001
4.5	X3sol	0.823	0.603	0.824	0.822	0.824	0.899	0.790	0.036	0.805	1.032	0.005	0.000

*The selected model.

4.3.3. Comparison of the Applicability Domain of $pEC_{50,NR[GFS]}$ Models

To select the best QSTR model for $pEC_{50,NR[GFS]}$ endpoint, the AD of 7 models (2-4 descriptor models in Tables 4.18 and 4.19) were compared. Models were used to predict $pEC_{50,NR[GFS]}$ values for an external set composed of 660 chemicals with no experimental $pEC_{50,NR[GFS]}$ values. Structural coverage of each model and number of chemicals predicted out of 660 were given in Table 4.20.

Table 4.20. Predictive performances of the generated $pEC_{50,NR[GFS]}$ models.

Division no	Model no	Number of chemicals within the AD of models (out of 660)	Structural coverage (%)
1	1.1	627	95.0
	1.2	622	94.2
2	***2.1**	**627**	**95.0**
3	3.1	582	88.2
4	4.1	569	86.2
	4.2	571	86.5
	4.3	572	86.7

*The selected model.

By comparing all the fit, internal validation and external validation parameters and structural coverage of the models (Table 4.20), model 2.1 was selected as the best model. Fit, internal validation and external validation parameters of model 2.1 were written in bold in Tables 4.18 - 4.19, respectively. The resulting mathematical QSTR equation is given in Eq 4.7. Numbers in parentheses show the 95% confidence intervals for variable coefficients in the equation.

$$pEC_{50,NR[GFS]} = -2.053\ (\pm 0.398)\ GATS1v + 0.378\ (\pm 0.060)\ DP07 + 0.683\ (\pm 0.515) \quad (4.7)$$

The selected model was one of the best among the selected models in terms of fit parameters with high $R^2 = 0.836$, $R^2_{adj} = 0.809$ and low $RMSE_{TR} = 0.620$ values. Among all the other models, the selected model had the highest value for the parameter $Q^2_{LOO} = 0.799$ which was both beyond the acceptable limit and comparable with the R^2 value indicating a stable model. Significantly low $R^2_{Yscr} = 0.026$ and $Q^2_{Yscr} = -0.051$ values guaranteed that the generated model was not a result of a chance correlation. The external performance of the model was verified by very high $R^2_{TEST} = 0.977$ and low $RMSE_{TEST} = 0.226$ values. In terms of other external validation parameters, Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{TEST}, the model had satisfying results higher than the literature threshold values. Also, the homogeneity of the test set division in terms of response distribution was demonstrated by the parameters Q^2_{F1} and Q^2_{F2} which had the same value (0.975 for both). The model was successful in

terms of all the other external validation parameters (Table 4.19). Additionally, the external performance of the model was also verified with Golbraikh and Tropsha's criteria which certified external power of the model. Finally, the model was found to have "good predictions" in terms of MAE criteria (Roy et al., 2016) with *MAE* (95% of the data) = 0.171, 3σ = 0.326 and training set range = 5.501.

Chemicals used for the QSTR modelling of $pEC_{50,NR[GFS]}$ together with their training/test set status in modelling, experimental and predicted $pEC_{50,NR[GFS]}$ values, hat and descriptor values are given in Table 4.21.

Table 4.21. Chemicals used for the QSTR modelling of $pEC_{50,NR[GFS]}$, their experimental and predicted $pEC_{50,NR[GFS]}$ values, hat values and descriptor values.

Name	Status	Exp. $pEC_{50,NR[GFS]}$ (mM)	Pred. $pEC_{50,NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07
1,2,4-trichlorobenzene	Training	0.690	0.835	0.030	0.565	3.468
1,4-dioxane	Training	-2.730	-1.285	0.041	1.024	0.355
1-butanol	Test	-1.810	-1.790	0.054	1.611	2.205
1-napthol	Training	0.470	0.369	0.018	0.761	3.300
1-octanol	Training	-0.260	0.141	0.083	1.557	7.020
1-propanol	Test	-2.180	-2.391	0.067	1.645	0.802
2,3,4,6-tetrachlorophenol	Test	1.600	1.399	0.046	0.378	3.944
2,3,4-trichlorophenol	Training	1.680	0.987	0.038	0.476	3.388
2,3-dichlorophenol	Test	0.660	0.489	0.035	0.574	2.602
2,4,5-trichlorophenol	Training	1.790	1.136	0.036	0.476	3.781
2,4,6-trichlorophenol	Training	0.380	1.067	0.037	0.476	3.598
2,4,6-triiodophenol	Training	1.820	1.437	0.042	0.411	4.226
2,4-dichlorophenol	Training	0.840	0.730	0.030	0.574	3.240
2,4-dimethylphenol	Training	0.040	-0.266	0.015	1.009	2.968
2,5-dichlorophenol	Training	0.780	0.780	0.030	0.574	3.371
2,6-dichlorophenol	Training	-0.210	0.541	0.034	0.574	2.739
2-bromophenol	Training	0.230	0.176	0.035	0.638	2.123
2-butanol	Training	-2.130	-2.306	0.063	1.611	0.841
2-buthoxyethanol	Training	-1.860	-1.558	0.049	1.579	2.646
2-chlorophenol	Training	-0.290	0.001	0.033	0.691	1.948
2-ethoxyethanol	Training	-2.540	-1.850	0.055	1.614	2.064
2-methoxyethanol	Training	-2.900	-2.394	0.067	1.646	0.799
2-methylphenol	Training	-0.360	-0.566	0.023	0.940	1.799
2-nitrophenol	Training	-0.640	0.161	0.019	0.788	2.898

Table 4.21. Continued.

Name	Status	Exp. $pEC_{50,NR[GFS]}$ (mM)	Pred. $pEC_{50,NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07
2-pentanol	Training	-1.780	-1.688	0.051	1.590	2.361
2-propanol	Test	-2.710	-2.634	0.074	1.645	0.158
3,4-dichlorophenol	Training	1.200	0.696	0.031	0.574	3.150
3,5-dichlorophenol	Training	1.390	0.686	0.031	0.574	3.123
3,5-dimethoxyphenol	Training	-0.460	0.186	0.015	1.067	4.479
3,5-dimethylphenol	Training	-0.270	-0.316	0.015	1.009	2.836
3,5-di-*tert*-butylphenol	Training	1.680	0.232	0.027	1.217	5.413
3-bromophenol	Training	0.580	0.350	0.030	0.638	2.583
3-chlorophenol	Training	0.070	0.168	0.028	0.691	2.389
3-methylphenol	Test	-0.410	-0.399	0.019	0.940	2.241
4-bromophenol	Test	0.310	0.518	0.027	0.638	3.027
4-chloro-3,5-dimethyl phenol	Training	1.030	0.153	0.015	0.870	3.321
4-chlorophenol	Training	-0.110	0.362	0.024	0.691	2.903
4-methyl-2-Nitrophenol	Training	0.010	0.284	0.013	0.883	3.737
4-methylphenol	Test	-0.070	-0.224	0.016	0.940	2.704
4-nitrophenol	Training	0.530	0.566	0.015	0.788	3.968
Acetone	Training	-2.420	-2.310	0.059	1.485	0.147
Acetonitrile	Training	-3.020	-1.808	0.047	1.232	0.102
Anilazine	Training	1.190	1.811	0.031	0.642	6.469
Aniline	Test	-1.050	-0.552	0.031	0.853	1.364
Benzene	Training	-1.240	-1.055	0.043	0.917	0.381
Benzoximate	Training	1.230	1.608	0.046	0.977	7.749
Bibenzyl	Training	1.220	1.456	0.035	0.931	7.097
Captafol	Training	1.710	1.557	0.028	0.789	6.594
Captan	Training	1.660	1.741	0.030	0.639	6.267

Table 4.21. Continued.

Name	Status	Exp. $pEC_{50,NR[GFS]}$ (mM)	Pred. $pEC_{50,NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07
Chlorobenzilate	Training	1.760	1.764	0.036	0.814	7.276
Chlorothalonil	Training	2.260	1.686	0.036	0.468	5.192
Chlorpyrifos	Training	1.770	1.443	0.030	0.882	6.798
Chlorpyrifos-methyl	Training	1.420	1.225	0.021	0.821	5.890
Chloroform	Training	-1.070	-0.281	0.070	0.541	0.387
Cyanophos	Training	0.730	1.359	0.034	0.965	7.027
Diazinon	Training	0.970	0.712	0.039	1.205	6.617
Dicofol	Training	2.180	2.094	0.039	0.611	7.048
Diethyl ether	Training	-2.430	-2.502	0.077	1.742	1.035
di-*n*-butylphthalate	Training	1.510	0.938	0.046	1.190	7.133
Dioxabenzofos	Training	0.430	0.593	0.014	0.913	4.718
Diphenyl	Training	0.910	0.882	0.016	0.837	5.069
Disulfoton	Training	0.850	0.215	0.019	1.142	4.962
Edifenphos	Training	1.200	1.868	0.044	0.856	7.781
EPN	Training	1.820	1.786	0.044	0.893	7.764
Ethanol	Training	-2.350	-2.799	0.083	1.709	0.071
Fenitrothion	Training	1.320	1.289	0.039	1.031	7.199
Fenobucarb	Training	-0.170	0.175	0.020	1.157	4.937
Fenthion	Test	1.350	1.182	0.033	1.018	6.846
gamma-BHC	Training	1.210	0.732	0.024	0.632	3.560
Iprofenfos	Training	0.380	0.866	0.038	1.152	6.738
Isobutanol	Training	-1.800	-2.314	0.063	1.611	0.820
Isoprothiolane	Training	0.060	0.995	0.033	1.076	6.666
Isoxathion	Test	1.330	1.808	0.060	1.004	8.424

Table 4.21. Continued.

Name	Status	Exp. $pEC_{50,NR[GFS]}$ (mM)	Pred. $pEC_{50,NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07
Methanol	Training	-2.700	-3.157	0.106	1.871	0.002
Methidathion	Training	0.270	0.826	0.018	0.959	5.585
Molinate	Training	-0.200	0.264	0.031	1.248	5.666
m-xylene	Training	0.080	-0.612	0.018	1.063	2.345
Naled	Test	1.010	0.709	0.014	0.879	4.840
Naphthalene	Training	0.280	0.165	0.018	0.805	3.001
Oxadiazon	Test	1.440	1.538	0.045	0.996	7.668
o-xylene	Training	0.220	-0.787	0.022	1.063	1.883
p,p-DDD	Test	1.790	1.955	0.036	0.685	7.081
Pentachlorophenol	Training	2.170	1.805	0.064	0.235	4.242
Phenthoate	Training	0.890	-0.793	0.031	1.394	3.665
Phenol	Training	-0.900	-0.569	0.032	0.847	1.287
Propanil	Training	0.330	1.357	0.024	0.810	6.179
Propaphos	Test	0.840	1.058	0.050	1.181	7.403
Propargite	Training	2.960	1.729	0.079	1.137	8.938
p-xylene	Test	-0.230	-0.417	0.015	1.063	2.863
Pyrazolate	Training	1.440	0.014	0.026	0.745	2.274
Pyridine	Training	-1.990	-1.105	0.043	0.938	0.364
Quintozene	Test	2.160	2.124	0.069	0.190	4.842
Salicylaldehyde	Training	1.090	0.006	0.020	0.817	2.644
Simetryn	Training	-0.490	0.758	0.034	1.150	6.442
tert-amyl alcohol	Test	-2.050	-2.260	0.061	1.590	0.850
tert-butanol	Training	-2.380	-2.532	0.069	1.611	0.243
Trifluralin	Training	1.600	1.168	0.048	1.138	7.459

The plot of predicted $pEC_{50,NR[GFS]}$ values from Eq 4.7 versus experimental $pEC_{50,NR[GFS]}$ values is given in Figure 4.12. An agreement between the experimental and predicted $pEC_{50,NR[GFS]}$ values is claimed as the data aligned uniformly along the optimal line.

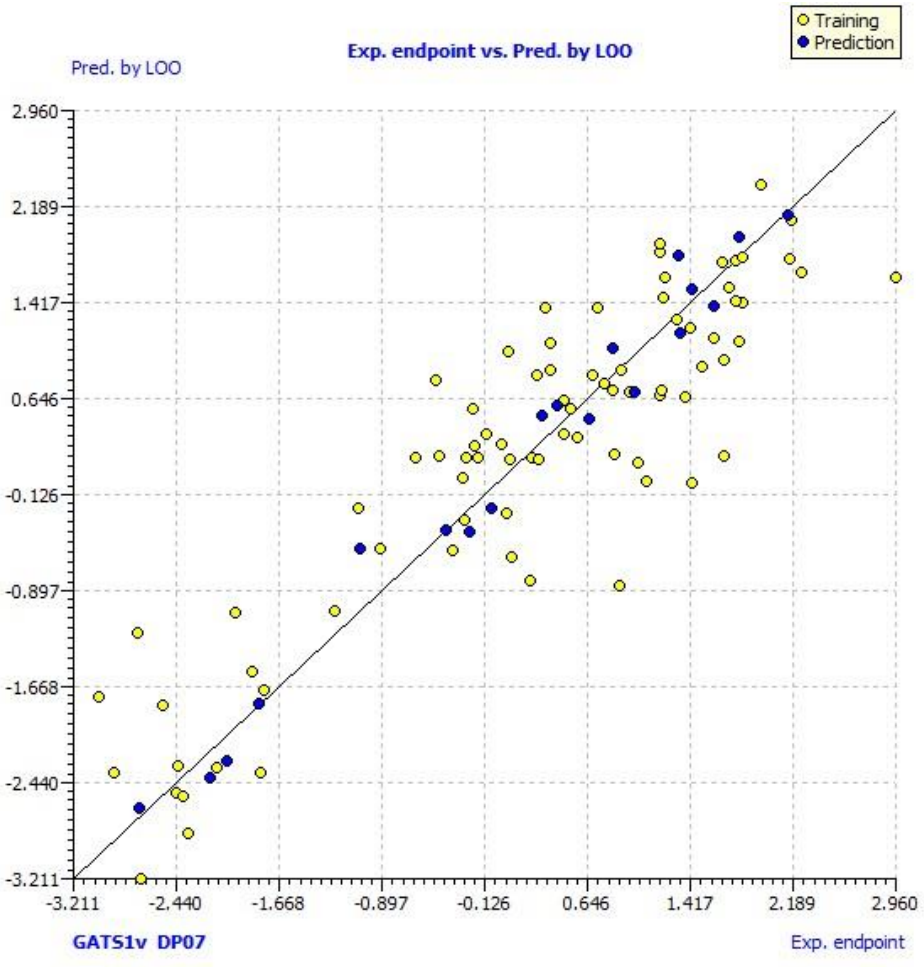


Figure 4.12. Plot of predicted $pEC_{50,NR[GFS]}$ values from Eq. 4.7 versus experimental $pEC_{50,NR[GFS]}$ values.

The AD of the Eq. 4.7 can be investigated with Williams plot (Figure 4.13). It is important to state that neither response nor structural outliers were present because all the training and test set compounds had residuals lower than ±2.5σ and none of them had hat values greater than the critical hat value of the model (h^* = 0.111), respectively.

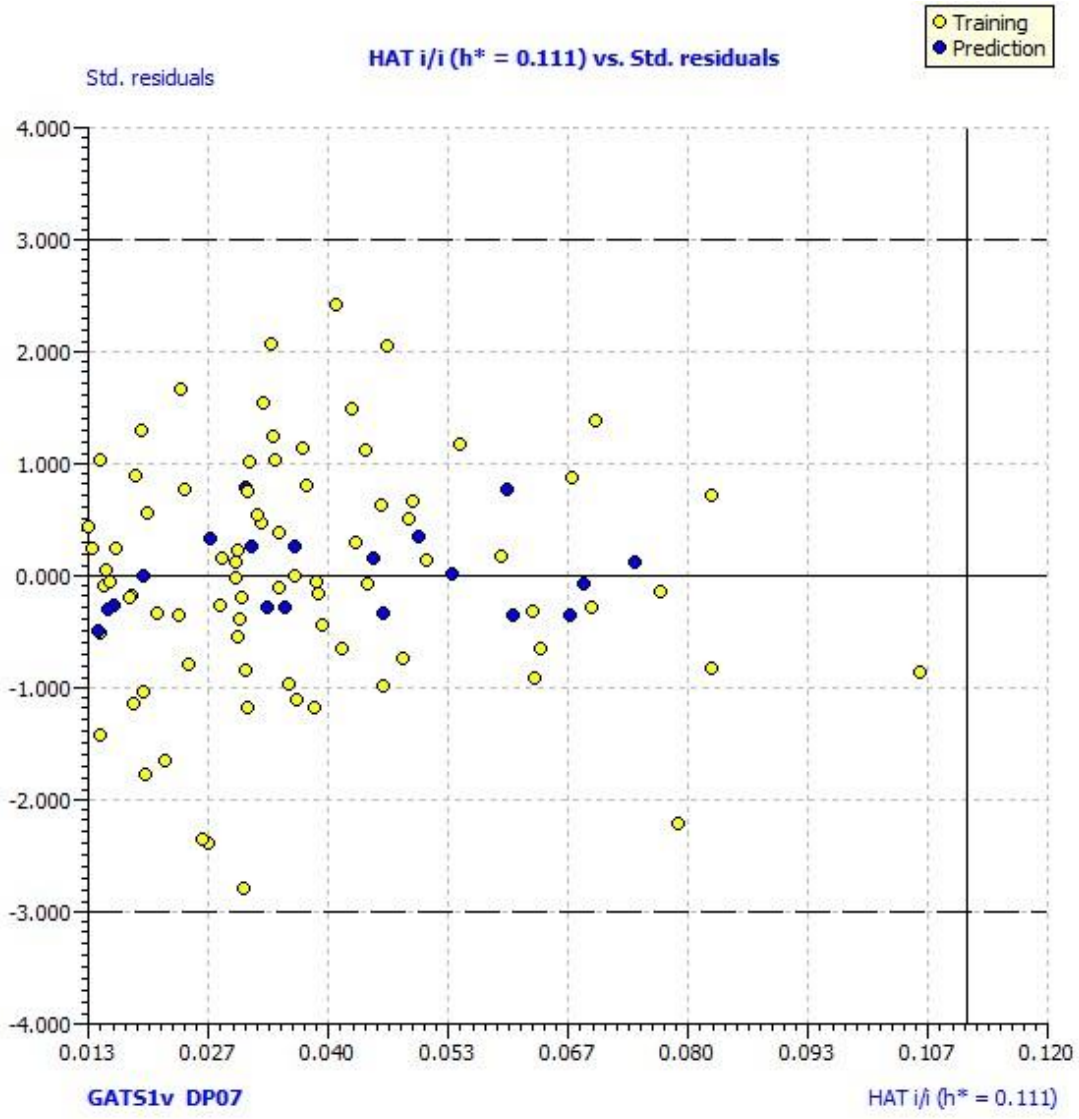


Figure 4.13. Williams plot of $pEC_{50,NR[GFS]}$ model.

The external predictive ability of the model was evaluated with an external set composed of 660 diverse chemicals. External set chemicals, calculated descriptor, predicted and experimental $pEC_{50,NR[GFS]}$ values are provided in Appendix C. Insubria graph, in which predicted $pEC_{50,NR[GFS]}$ values for training, test and external set chemicals versus their hat values can be investigated, is given in Figure 4.14.

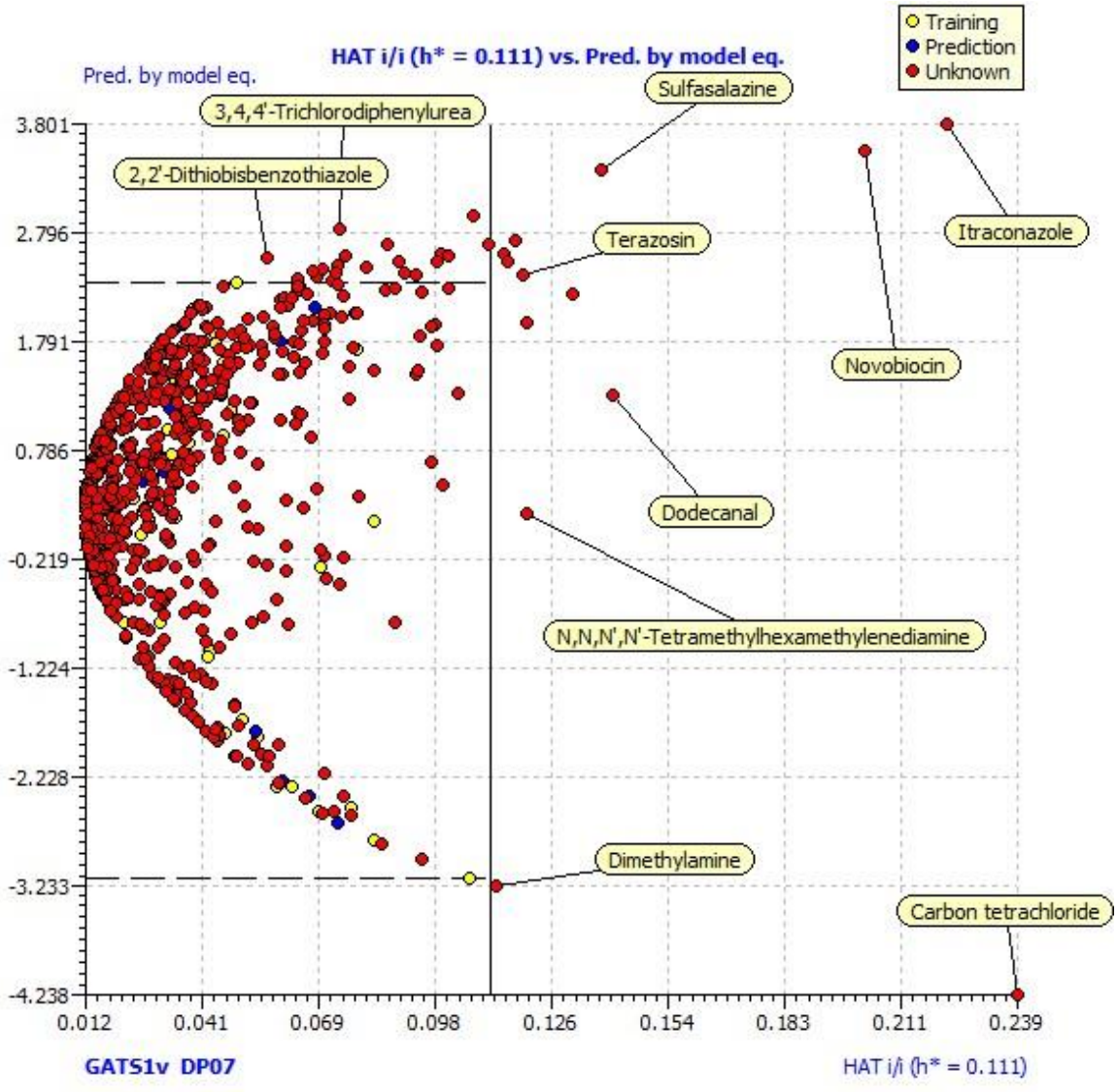


Figure 4.14. Insubria graph of $pEC_{50,NR[GFS]}$ model. Predicted $pEC_{50,NR[GFS]}$ values for training, test and external (660 chemicals) set chemicals by Eq. 4.7 versus their hat values.

The predicted $pEC_{50,NR[GFS]}$ values were used to screen the most and the least toxic chemicals among the external set chemicals (Table 4.22). Methotrexate was found to be the most cytotoxic chemical with the highest $pEC_{50,NR[GFS]}$ value. Methotrexate (4-amino-10-methyl-folic acid) is an antineoplastic agent used in chemotherapy to inhibit the activity of dihydrofolate reductase enzyme (Roig et al., 2014).

Table 4.22. The most and the least cytotoxic chemicals from the external set predicted by Eq. 4.7.

	Name	Chemical class	$pEC_{50,NR[GFS]}$ pred. by Eq. 4.7 (mM)
Most Toxic	Methotrexate	Pharmaceutical, Antineoplastic Agents	2.335
	Tetrabromobisphenol A	Diphenyl alkane,dihydroxy,bromo	2.310
	Cephalexin	Pharmaceutical, Anti-Bacterial Agents	2.308
	Oxytetracycline	Pharmaceutical, Anti-bacterial agents	2.307
	Cyclosulfamuron	Pesticide, Herbicide	2.291
	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	Diphenyl alkane,alkoxy-alcohol	2.289
	Cefurexime axetil	Pharmaceutical, Anti-Bacterial Agent	2.280
	Indomethacin	Pharmaceutical, Nonsteroidal Antiinflammatory Drugs	2.279
	Bensulfuron-methyl	Pesticide, Herbicide	2.250
	Acebutolol	Pharmaceutical, Beta-Adrenergic Receptor Antagonists	2.246
Least Toxic	*n*,*n*-Dimethylhydrazine	Hydrazine	-2.973
	Methylhydrazine	Hydrazine	-2.836
	Diethylamine	Amine-NH	-2.575
	t-butylamine	Amine, butyl	-2.546
	Dimethylnitrosamine	Nitrosamine, methyl	-2.541
	Dimethylformamide	Formamide, methyl	-2.407
	Orthoformic acid trimethyl ester	Acid, ester	-2.396
	tert-Butylhydroperoxide	Hydroperoxide	-2.275
	Triethylamine	Amine-N	-2.181
	3-pentanol	Alcohol	-2.111

The reported concentrations of methotrexate in different water sources range from about 1 μg/L in sewage effluent to less than 6.25 ng/L in river and potable water (Aherne and English, 1985). Similar to high cytotoxic potential reported in the present study, ecotoxic potential of methotrexate was also indicated in a study by Henschel and co-workers (1997) depending on its impact on the proliferation rate of certain cell cultures and its teratogenic effects. As a result, it can be suggested that the control of methotrexate in terms of usage/disposal would be beneficial from an environmental risk perspective.

Additionally, the most toxic chemicals reported in Table 4.22 were mainly from pharmaceuticals (6 of 10 chemicals were pharmaceutical) in line with the $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[PLHC-1]}$ predictions reported in the present study. Hence, we suggest the need for the assessment and screening for this type of chemicals.

Molecular descriptors appeared in Eq. 4.7 are given in Table 4.23 together with the standardized coefficients which reflect the relative importance of each descriptor in the model.

Table 4.23. Descriptors appeared in the $pEC_{50,NR[GFS]}$ model.

Descriptor	Meaning of descriptor	Type	Standardized coefficient
DP07	Molecular profile no. 7	Randić molecular profiles	0.628
GATS1v	Geary autocorrelation of lag 1 weighted by van der Waals volume	2D autocorrelations	-0.517

DP07, the most important descriptor in the model, belongs to the Randić molecular profiles. Molecular profile descriptors (^{k}D) derived from powers of the geometric distance matrix (G) (Todeschini and Consonni, 2008). G is a A×A, square symmetry matrix where each entry is the Euclidian distance between two atoms (distance between atoms s and t; r_{st}) (Todeschini and Consonni, 2008).

$$G \equiv \begin{matrix} 0 & r_{12} & \cdots & r_{1A} \\ r_{21} & 0 & \cdots & r_{2A} \\ \cdots & \cdots & \cdots & \cdots \\ r_{A1} & r_{A2} & \cdots & 0 \end{matrix} \qquad (4.8)$$

Molecular profile descriptors are calculated by averaging sums of k^{th} power of G matrix entries and by normalizing with the k! factor. A general definition of ${}^{k}D$ is given in Eq 4.9.

$$ {}^{k}D = \frac{1}{k!} \frac{\sum_{i=1}^{A} \sum_{j=1}^{A} r_{ij}^{k}}{A} \tag{4.9} $$

where r_{ij}^{k} is the k^{th} power of i-j entry of Geometry matrix (G) and A is the number of atoms (Randić, 1995; Randić and Razinger, 1995 as referred in Todeschini and Consonni, 2008).

The other descriptor GATS1v, which is a 2D autocorrelations descriptor, negatively correlated with the endpoint. For a mechanistic approach, it might be useful to state that GATS1v descriptor values tend to decrease with the addition of chlorine substituents to phenolic compounds (which covers most of the chemicals in dataset) hence higher toxicity values. The same situation was valid for methyl substituents, as well. For chemicals with aliphatic chains, on the other hand, longer chains le d to higher GATS1v values thus lower toxicity values.

GATS1v descriptors have been used in previous studies for modeling of antifungal activity (Duchowicz et al., 2007), modeling of certain enzymatic activities (Amin et al., 2016) and for QSTR modeling of toxicity of esters to *Daphnia magna* by Papa and co-workers (2005). Moreover, Tuğcu and co-workers (2012) have also found a positive relationship between another 2D autocorrelations descriptor, GATS3p, and toxicity of pharmaceuticals to fish.

4.3.4. Correlation between $pEC_{50,NR[GFS]}$ and log K_{ow}

The log K_{ow} values ranged from -0.43 for 2-methoxyethanol to 5.89 for dicofol in $pEC_{50,NR[GFS]}$ dataset. The correlation analysis revealed that there is a strong correlation between log K_{ow} and $pEC_{50,NR[GFS]}$ values for all compounds (n = 100; r = 0.865; p < 0.01). The plot of observed pLC_{50} versus $pEC_{50,NR[GFS]}$ values is provided in Figure 4.15. There was only one compound, pyrazolate, having a weak relationship with log K_{ow} and $pEC_{50,NR[GFS]}$ and it is indicated in Figure 4.15 with red marker. The strong correlation suggests that log K_{ow} might be a useful property to predict neutral red uptake dependent cytotoxicity of structurally similar chemicals to GFS cell line at least for a primary assessment/prioritization purpose.

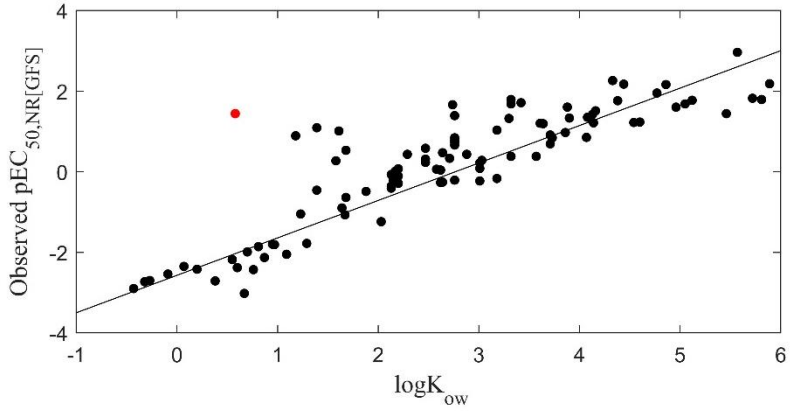


Figure 4.15. Plot of observed $pEC_{50,NR[GFS]}$ values versus log K_{ow} values.

4.3.5. Correlation of Experimental *in vivo* pLC_{50} with Predicted *in vitro* $pEC_{50,NR[GFS]}$

To assess the suitability of the proposed model as an alternative of *in vivo* toxicity testing, the relationship between the predicted $pEC_{50,NR[GFS]}$ values and the fish toxicity (pLC_{50}) data compiled from the literature were evaluated. The experimental pLC_{50} and predicted $pEC_{50,NR[GFS]}$ values for external set chemicals are given in Appendix C.

A strong correlation (n = 189; r = 0.574; p < 0.01) was found between the predicted *in vitro* $pEC_{50,NR[GFS]}$ and experimental *in vivo* pLC_{50} values (Figure 4.16). Of the 230 external set chemicals, a weak relationship between the predicted $pEC_{50,NR[GFS]}$ and the experimental pLC_{50} values was observed for 41 chemicals. Phenols and benzene derivatives in the data set were the predominant groups having a weak relationship yet chemicals from various classes were also present.

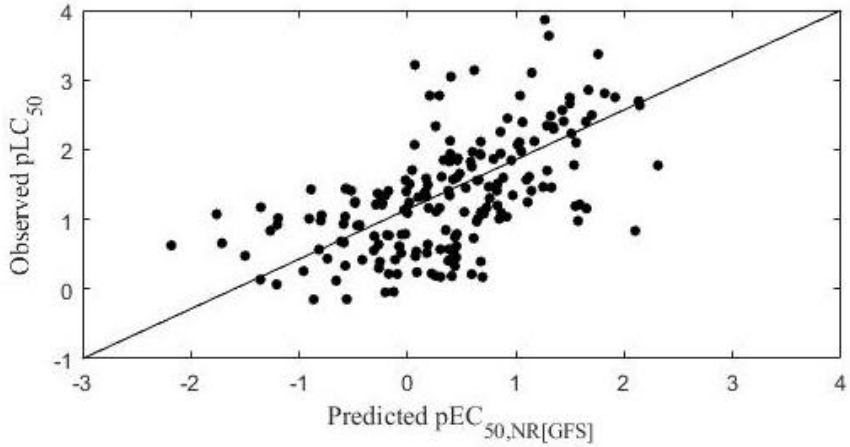


Figure 4.16. Plot of observed pLC_{50} versus predicted $pEC_{50,NR[GFS]}$ values by Eq. 4.7.

4.4. QSTR Modelling for $pEC_{50,NR[FHM]}$

4.4.1. Dataset

Dataset compounds together with their physicochemical properties, chemical classes, and experimental $pEC_{50,NR[FHM]}$ values are listed in Table 4.24.

Table 4.24. The name, CAS number, log K_{ow}, molecular weight and experimental $pEC_{50,NR[FHM]}$ values of chemicals in the dataset.

ID	Compound Name	CAS number	Molecular Weight*	$pEC_{50,NR[FHM]}$ (mM)	log K_{ow}**
1	Acetaldehyde	75-07-0	44.053	-1.079	-0.57
2	Acetic acid	64-19-7	60.052	-1.690	-0.31
3	Acetone	67-64-1	58.080	-2.772	0.20
4	Acetonitrile	75-05-8	41.053	-2.956	0.67
5	Benzoic acid	65-85-0	122.123	-1.362	1.59
6	Butyric acid	107-92-6	88.106	-1.716	0.76
7	Chloral Hydrate	302-17-0	165.403	-1.114	1.38
8	Citric acid	77-92-9	192.123	-1.146	-1.68
9	Cumene hydroperoxide	80-15-9	152.193	0.086	2.03
10	Diacetone alcohol	123-42-2	116.16	-2.430	0.33
11	Dichloromethane	75-09-2	84.933	-2.260	1.01
12	Diethanolamine	111-42-2	105.137	-1.342	-1.17
13	Epichlorohydrin	106-89-8	92.525	0.108	0.53
14	Ethanol	64-17-5	46.069	-2.944	0.07
15	Ethanolamine	141-43-5	61.084	-0.431	-1.17
16	Ethyl acetate	141-78-6	88.106	-1.934	0.29
17	Ethyl acetoacetate	141-97-9	130.143	-0.792	1.00
18	Furfural	98-01-1	100.117	-0.881	-0.44
19	Furfuryl alcohol	98-00-0	102.133	-1.869	-0.13
20	Isobutanol	78-83-1	74.123	-2.170	0.95
21	Isopropanol	67-63-0	60.096	-2.621	0.38
22	Morpholine	110-91-8	83.090	-1.964	-0.56***
23	*n*-butanol	71-36-3	74.123	-2.079	0.97
24	*n*-butanone	78-93-3	72.107	-2.297	0.86
25	*n*-butylamine	109-73-9	73.139	-0.792	0.59
26	*n*-heptanol	111-70-6	116.204	-0.591	2.22
27	*n*-octanol	111-87-5	130.231	-0.301	2.64
28	*n*-pentanol	71-41-0	88.150	-1.633	1.39
29	*n*-propanol	71-23-8	60.096	-2.477	0.55
30	Oxalic acid	144-62-7	90.034	-1.230	-0.53

Table 4.24. Continued.

ID	Compound Name	CAS number	Molecular Weight*	$pEC_{50,NR[FHM]}$ (mM)	log K_{ow}**
31	Phenol	108-95-2	94.113	-0.892	1.64
32	Propyl acetate	109-60-4	102.133	-1.778	0.78
33	Tetrahydrofuran	109-99-9	72.107	-2.326	0.40

*MW calculated by SPARTAN 10, ***n*-octanol-water partition coefficients calculated by SPARTAN 10, ***calculated by ECOSAR v1.11

4.4.2. Model Development

Dataset for $pEC_{50,NR[FHM]}$ model included chemicals from distinct classes. For such a diverse dataset, it was rather complicated to find the best combination of the descriptors meeting all the fit, internal and external validation criteria. Therefore, although models were generated for various divisions, only two random divisions among those provided models in accordance with all the statistical criteria. Three-descriptor models were generated for these divisions. Table 4.25 includes test set compound's ID and the ratio of the number of test set to the number of training set compounds for these divisions.

Fit, internal validation and external validation parameters for the generated two models were given in Table 4.26 and Table 4.27, respectively. Two generated models were acceptable in terms of both fit and internal validation parameters with $R^2 > 0.5$, $Q^2_{LOO} > 0.5$ values indicating stability and robustness of the models, low $RMSE_{TR}$ and $RMSE_{TEST}$ values, and external validation parameters with $R^2_{TEST} > 0.5$, $Q^2_{Fn} > 0.7$ and $CCC_{TEST} > 0.8$ for both. However, as stated above, it was not possible to generate models both covering all of the chemicals in the data set and meeting with all the statistical requirements.

Table 4.25. Test set compounds for two divisions used in the QSTR modelling of $pEC_{50,NR[FHM]}$.

Division no	n_{TEST}/n_{TR}	Test Set Compounds*
1	6/29	7, 12, 13, 20, 25, 30
2	7/28	7, 17, 19, 22, 32, 33, 35

*Compound numbers refer to the ID numbers given in Table 4.24.

Table 4.26. Fit and internal validation parameters of the generated $pEC_{50,NR[FHM]}$ models.

Model	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
***1**	**X5v R3m+ GATS2v**	**0.627**	**0.580**	**0.538**	**0.770**	**13.417**	**0.502**	**0.621**	**0.689**
2	MATS2v R3m+ F03[C-C]	0.631	0.585	0.530	0.774	13.684	0.529	0.599	0.705

*The selected model.

Table 4.27. External validation parameters of the generated $pEC_{50,NR[FHM]}$ models.

Model	Descriptors	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2–r_0'^2)/r^2$	$(r^2–r_0^2)/r^2$
***1**	**X5v R3m+ GATS2v**	**0.893**	**0.155**	**0.889**	**0.884**	**0.969**	**0.938**	**0.844**	**0.097**	**0.967**	**1.026**	**0.013**	**0.000**
2	MATS2v R3m+ F03[C-C]	0.858	0.204	0.837	0.812	0.946	0.918	0.742	0.137	0.987	1.000	0.004	0.054

*The selected model.

4.4.3. Comparison of the Applicability Domain of $pEC_{50,NR[FHM]}$ Models

To select the best model, two QSTR models were used to predict $pEC_{50,NR[FHM]}$ values of an external set of 230 chemicals from various classes. Structural coverage of each model together with the number of compounds within the AD of $pEC_{50,NR[FHM]}$ model are given in Table 4.28.

Table 4.28. Predictive performances of the generated $pEC_{50,NR[FHM]}$ models.

Model	Number of chemicals within the AD of $pEC_{50,NR[FHM]}$ model (out of 230)	Structural coverage (%)
***1**	**205**	**89.1**
2	198	86.1

*The selected model.

On the basis of structural coverage, model 1 from Tables 4.26 and 4.28 was selected as the best model. Fit, internal and external validation parameters and structural coverage of model 1 is written in bold through Tables 4.26- 4.28, respectively. Model 1 has the following mathematical equation where the numbers in parenthesis indicate the 95% confidence intervals of the coefficients:

$$pEC_{50,NR[FHM]} = 3.301\ (\pm 1.463)\ X5v + 5.005\ (\pm 2.285)\ R3m+ + 1.436\ (\pm 1.175)\ GATS2v - 3.971\ (\pm 1.463) \quad (4.10)$$

The selected model was acceptable in terms of fit parameters with $R^2 = 0.627$, $R^2_{adj} = 0.580$ and $RMSE_{TR} = 0.538$. Its stability was demonstrated by the parameter $Q^2_{LOO} = 0.502$ but it is important to state that the value of Q^2_{LOO} parameter fulfilled just the acceptance criterion and was not better. Any possibility of chance correlation was discarded considering the significantly low $R^2_{Yscr} = 0.112$ and $Q^2_{Yscr} = -0.248$ values. As can be seen in Table 4.26, even though the model did not show a high performance in terms of the fit and internal validation metrics, all the presented results satisfied acceptance criteria. Moreover, the model showed a much better performance in terms of external validation parameters: high $R^2_{TEST} = 0.893$ and low $RMSE_{TEST} = 0.538$ values indicated the predictive power of the model. Also, the predictive ability of the model was further demonstrated with the other external validation parameters, Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, and CCC_{TEST} as the values were much higher than the corresponding literature threshold values. Closer values of $Q^2_{F1} = 0.889$ and $Q^2_{F2} = 0.884$ promoted a homogeneous response distribution for the test set chemicals. Following the Golbraikh and Tropsha's criteria, the model performed well in terms of all the expected conditions. To finalize Eq. 4.10 as $pEC_{50,NR[FHM]}$ model, it was subjected to the MAE-based criteria (Roy et al., 2016). It satisfied

the criteria with *MAE* (95% of the data) = 0.127, 3σ = 0.294 and training set range = 3.064 such that the model predictions qualified as being good

Chemicals used for the QSTR modelling of $pEC_{50,NR[FHM]}$ together with their training/test set status in modelling, experimental and predicted $pEC_{50,NR[FHM]}$ values, hat and descriptor values are given in Table 4.29.

Table 4.29. Chemicals used for the QSTR modellling of $pEC_{50,NR[FHM]}$, their experimental and predicted $pEC_{50,NR[FHM]}$ values, hat values and descriptor values.

Chemicals	Status	Exp. $pEC_{50,NR[FHM]}$ (mM)	Pred. $pEC_{50,NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h*=0.429)	X5v	R3m+	GATS2v
Acetaldehyde	Training	-1.079	-2.315	0.080	0.000	0.022	1.077
Acetic acid	Training	-1.690	-2.376	0.091	0.000	0.021	1.038
Acetone	Training	-2.772	-2.575	0.148	0.000	0.017	0.913
Acetonitrile	Training	-2.956	-2.608	0.164	0.000	0.020	0.880
Benzoic acid	Training	-1.362	-0.903	0.118	0.330	0.077	1.110
Butyric acid	Test	-1.716	-1.872	0.071	0.000	0.047	1.298
Chloral Hydrate	Training	-1.114	-0.985	0.491	0.000	0.424	0.602
Citric acid	Training	-1.146	-0.992	0.065	0.236	0.095	1.201
Cumene hydroperoxide	Training	0.086	-0.642	0.354	0.510	0.058	0.944
Diacetone alcohol	Training	-2.430	-2.332	0.092	0.000	0.042	0.995
Dichloromethane	Test	-2.260	-2.264	0.078	0.000	0.000	1.189
Diethanolamine	Test	-1.342	-1.285	0.089	0.112	0.056	1.418
Epichlorohydrin	Training	0.108	-0.058	0.520	0.000	0.451	1.153
Ethanol	Training	-2.945	-2.011	0.081	0.000	0.015	1.313
Ethanolamine	Training	-0.431	-1.439	0.168	0.000	0.073	1.509
Ethyl acetate	Training	-1.935	-2.212	0.086	0.000	0.073	0.971
Ethylacetoacetate	Training	-0.792	-1.302	0.044	0.164	0.070	1.238
Ethylenediamine	Training	-1.716	-1.485	0.193	0.000	0.051	1.554
Furfural	Test	-0.881	-0.976	0.062	0.172	0.143	1.192
Furfuryl alcohol	Training	-1.869	-1.057	0.061	0.220	0.076	1.259
Isobutanol	Training	-2.170	-1.901	0.056	0.000	0.063	1.222
Isopropanol	Training	-2.621	-2.154	0.071	0.000	0.012	1.224
Morpholine	Training	-1.964	-1.620	0.043	0.136	0.060	1.116
n-butanone	Training	-2.297	-2.049	0.058	0.000	0.076	1.074
n-butylamine	Training	-0.792	-1.792	0.124	0.000	0.022	1.441

Table 4.29. Continued.

Chemicals	Status	Exp. $pEC_{50,NR[FHM]}$ (mM)	Pred. $pEC_{50,NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h*=0.429)	X5v	R3m+	GATS2v
n-heptanol	Training	-0.591	-0.556	0.188	0.381	0.015	1.450
n-octanol	Training	-0.301	-0.129	0.304	0.506	0.016	1.457
n-pentanol	Test	-1.634	-1.468	0.093	0.112	0.016	1.430
n-propanol	Training	-2.477	-1.756	0.095	0.000	0.047	1.379
Oxalic acid	Training	-1.230	-1.203	0.118	0.000	0.235	1.109
Phenol	Training	-0.892	-1.246	0.065	0.242	0.045	1.185
Propyl acetate	Training	-1.778	-1.595	0.047	0.144	0.068	1.087
Tetrahydrofuran	Training	-2.326	-2.195	0.073	0.000	0.008	1.209

The linear relationship between predicted $pEC_{50,NR[FHM]}$ values from Eq. 4.10 versus experimental values $pEC_{50,NR[FHM]}$ is shown in Figure 4.17. It can be seen from Figure 4.17 that even though most of the chemicals showed a good alignment along the optimal line, some of them showed deviations as expected due to structural variance mentioned before. Nevertheless, a uniform distribution was still present hence the fit of the model was further supported along with statistical fit parameters.

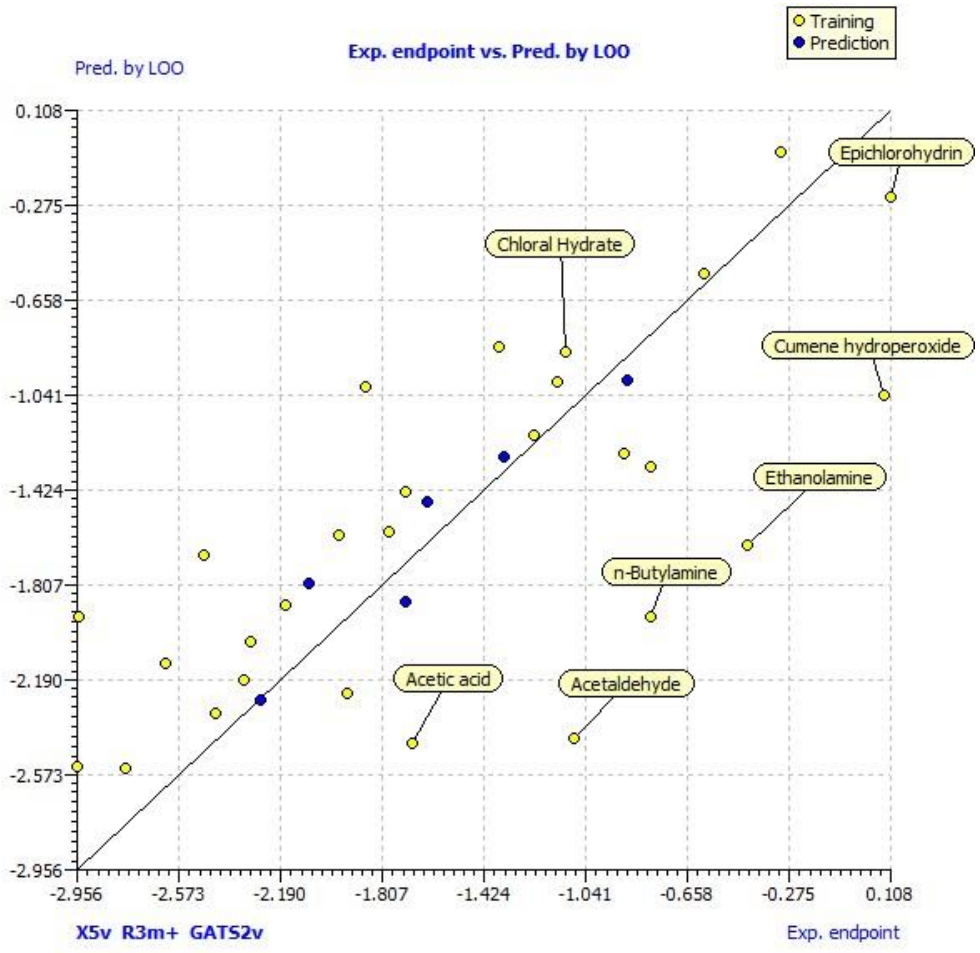


Figure 4.17. Plot of predicted $pEC_{50,NR[FHM]}$ values from Eq. 4.10 versus experimental $pEC_{50,NR[FHM]}$ values.

Figure 4.18 depicts Williams plot for the $pEC_{50,NR[FHM]}$ model with which any outlier from structural and response spaces can be detected. As all the chemicals in the training and test sets had residuals lower than $\pm 2.5\sigma$ thus there were no response outliers. However, as can be seen from the Figure 4.18 two structural outliers from the training set were present. Epichlorohydrin, one of the outliers, is a cyclic ether and structurally different from the rest of the data set. Dataset included some other cyclic chemicals but there was not another ether. Moreover, epichlorohydrin had the highest

R3m+ descriptor value among the dataset chemicals. The other outlier, chloral hydrate, is a chemical with hydroxide and chlorine substituents by which it differs from the rest of the dataset. It also had the lowest GATS2v and the highest R3m+ (after epichlorohydrin) values compared to other data set chemicals. As these two chemicals had far different structures compared to the other dataset chemicals, it is acceptable that they are out of the structural AD of the model. Nevertheless, both chemicals were within the response range with standardized residuals lower than $\pm 1.0\sigma$. Also for all the remaining chemicals in the data set, hat values were lower than the critical hat value of the model (h^* = 0.429).

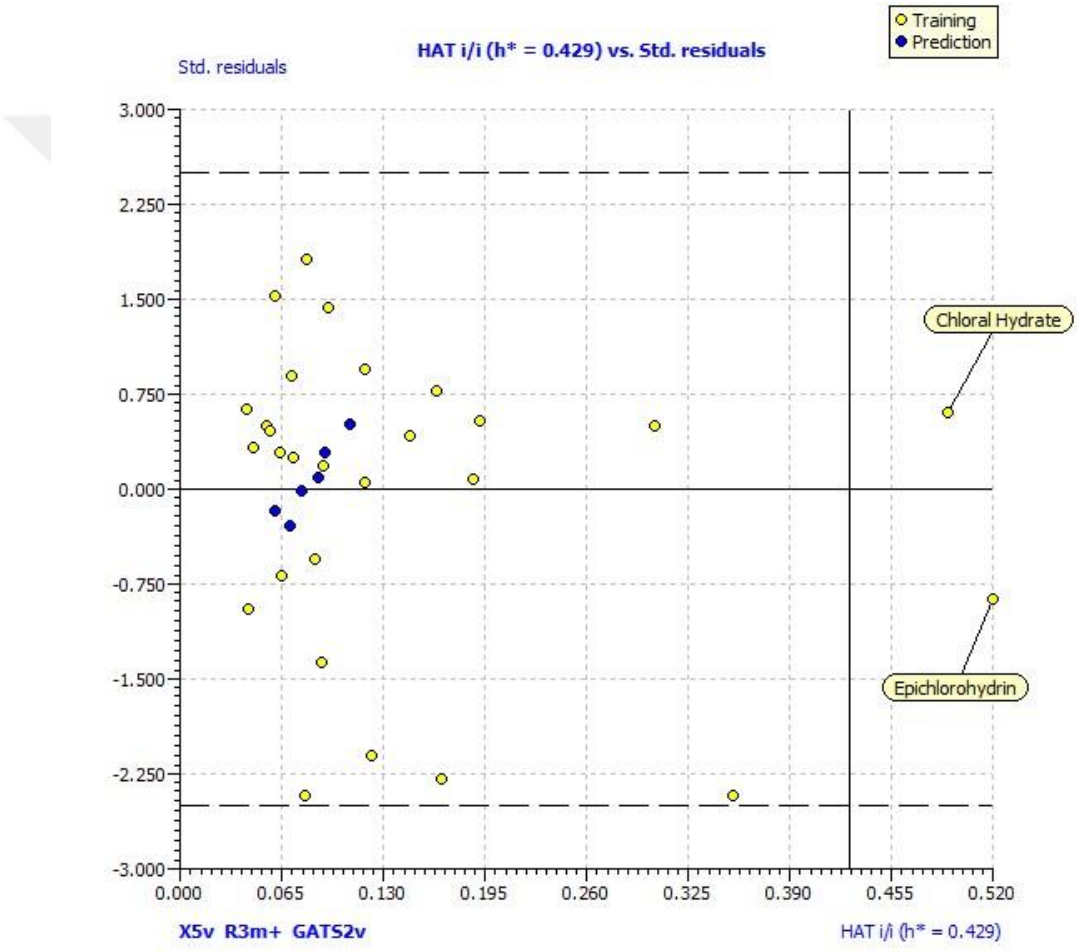


Figure 4.18. Williams plot of $pEC_{50,NR[FHM]}$ model.

The external predictive ability of the $pEC_{50,NR[FHM]}$ model was evaluated with an external set composed of 230 chemicals. External set chemicals, calculated descriptor, predicted and experimental $pEC_{50,NR[FHM]}$ values are provided in Appendix D. Of the 230 external set chemicals, the model was able to predict $pEC_{50,NR[FHM]}$ values for 205 chemicals. Insubria graph of the model where predicted

$pEC_{50,NR[FHM]}$ values for training, test and external set chemicals by Eq. 4.10 versus their hat values are plotted showing the structural coverage of the model, is given in Figure 4.19.

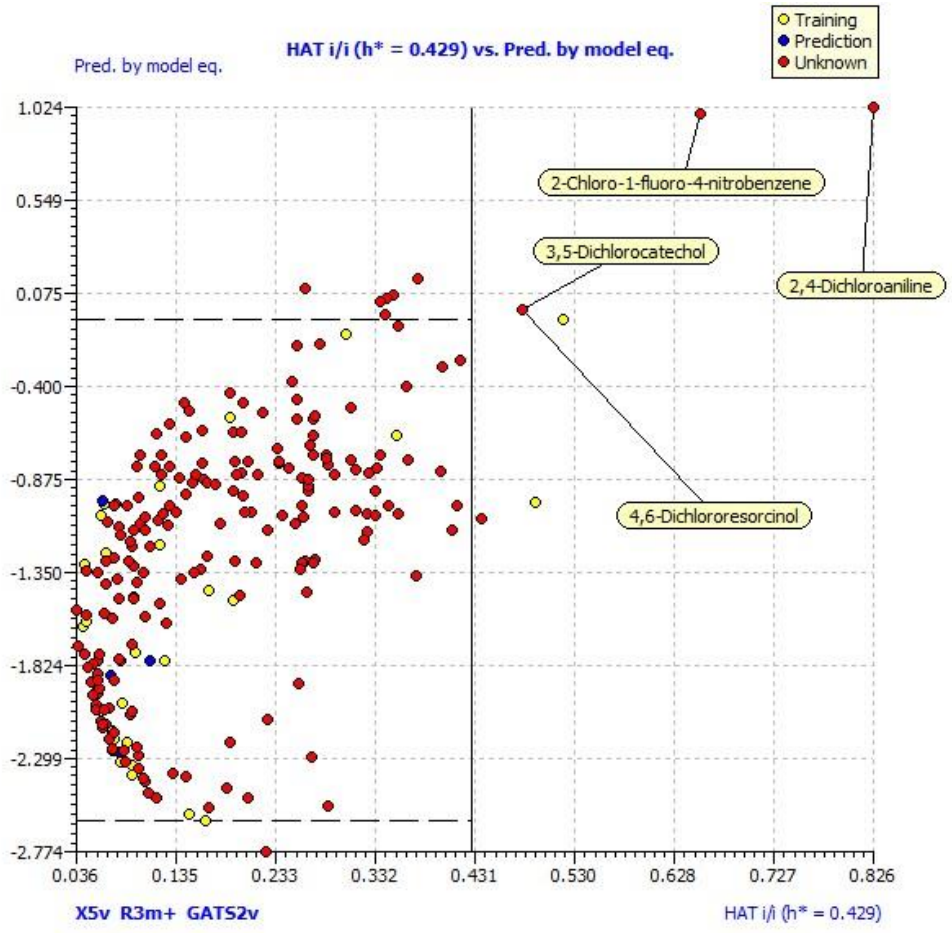


Figure 4.19. Insubria graph of $pEC_{50,NR[FHM]}$ model. Predicted $pEC_{50,NR[FHM]}$ values for training, test and external (230 chemicals) set chemicals from Eq. 4.10 versus their hat values.

As can be seen from the Insubria graph, 9 out of 25 chemicals which are dimethylamine; alpha-chloro-4-nitrotoluene; perfluorooctanoic acid; bromobenzene; alpha,alpha-dichlorotoluene, chlorohydroquinone; 1-cyclohexene-1-carbonitrile, hexamethylene diacrylate; benzyl-chloroformiate that cannot be predicted by the model were still within the structural AD of the model (with hat values less than the critical hat value of the model). However, they were out of the response range of the model thus their predcitions cannot be reliable.

Ten chemicals with the most and the least $pEC_{50,NR[FHM]}$ values predicted by Eq. 4.10 were screened and given in Table 4.30. The most cytotoxic chemical from the external set was found as 2-chlorobenzyl chloride with the highest predicted $pEC_{50,NR[FHM]}$ (-0.089 as mM) value.

Table 4.30. The most and the least cytotoxic chemicals from the external set with predicted by Eq. 4.10.

	Name	Chemical class	$pEC_{50,NR[FHM]}$ pred. by Eq. 4.10 (mM)
Most Toxic	2-chlorobenzyl chloride	Benzyl chloride,chloro	-0.089
	Octanoic acid	Acid	-0.179
	2-chloronitrobenzene	Nitrobenzene,chloro	-0.186
	1,3-dibromopropane	Alkane,bromo	-0.264
	Glyoxylic acid	Acid-aldehyde	-0.293
	o-chlorobenzonitrile	Benzonitrile, chloro	-0.374
	2,6-dichlorobenzonitrile	Pesticide, Herbicide	-0.394
	Octanedinitrile	Dinitrile	-0.427
	3-chloronitrobenzene	Nitrobenzene,chloro	-0.463
	Dibutylamine	Amine, butyl	-0.478
Least Toxic	1,1-dichloroethane	Alkane,chloro	-2.542
	Methyl methacrylate	Ester, acrylate	-2.533
	Methyl acrylate	Ester,a,b-unsaturated-acrylate	-2.498
	Methylhydrazine	Hydrazine	-2.493
	t-butylamine	Amine, butyl	-2.473
	Methacrylonitrile	Nitrile,a,b-unsaturated	-2.442
	Dimethyl disulphide	Disulfide	-2.412
	Thiourea	Thiourea	-2.399
	Pivalic acid	Acid	-2.388
	Methacrylic acid	Acid,a,b-unsaturated	-2.369

2-chlorobenzyl chloride is used as an intermediate for the production of certain agrochemicals. It is produced about 1,000 tonnes per year in Germany, Japan and Belgium (OECD, 2003). It is harmful for the environment as it may lead to long-term adverse effects in the aquatic organisms. The reported acute toxicity, LC_{50} (96 hr) of 2-chlorobenzyl chloride to different fish species were 0.27 mg/L, 0.5-0.71 mg/L and 0.71-0.96 mg/L for *Oryzias latipes, Danio rerio* and *Pimephales promelas*, respectively. Along with the previously reported toxic effects, cytotoxic potency of 2-chlorobenzyl chloride reported in the present study is important as it may provide a further emphasis on this chemical.

Molecular descriptors appeared in Eq. 4.10 are given in Table 4.31 with their standardized coefficients. Standardized coefficients were used to determine the relative importance of a descriptor in the model.

Table 4.31. Descriptors appeared in the $pEC_{50,NR[FHM]}$ model.

Descriptor	Meaning of descriptor	Type	Standardized coefficients
R3m+	R maximal autocorrelation of lag 3 / weighted by mass	GETAWAY descriptors	0.614
X5v	Valence connectivity index of order 5	Connectivity indices	0.595
GATS2v	Geary autocorrelation of lag 2 weighted by van der Waals volume	2D autocorrelations	0.342

R3m+ was the most important descriptor in Eq. 4.10 and it positively contributed to the $pEC_{50,NR[FHM]}$ endpoint. It belongs to Geometry, Topology and Atom-Weight Assembly (GETAWAY) descriptors which are originated from Molecular Influence Matrix (MIM) (Todeschini and Consonni, 2008). MIM consists of a matrix of Cartesian coordinates x, y, z of selected atoms in a molecule which is denoted by M. The general definition of MIM (denoted by H) is the following equation:

$$H=M\times(M^T\times M)^{-1}\times M^T \quad (4.11)$$

GETAWAY descriptors and specifically R autocorrelation descriptors have been used in previous ecotoxicology studies by Önlü and Sacan (2017) in the modeling of toxicity of industrial

chemicals and pharmaceuticals on algae and Papa and co-workers (2005) in the modelling of toxicity of esters to the planktonic crustacean. Also, Caballero and Fernandez (2006) used GETAWAY descriptors for antifungal activity modeling.

The other descriptor in the $pEC_{50,NR[FHM]}$ model, X5v, was a valence molecular connectivity index. Molecular connectivity indices, in general, encodes information about size, cyclization, branching, unsaturation and heteroatom content of molecules (Randić, 1975). The latter, heteroatom content, is embedded in an index that is called as molecular connectivity index ($^1X^v$) and developed by Kier and Hall, 1981.

To begin with, a simple molecular connectivity index, each atom (except hydrogen) in a molecule is assigned with a number depending on adjacency. Assigned numbers, called as δ values provide information on the number of nonhydrogenic-bonded atom pairs without giving information about elements of the pair or type of bonding present.

For calculation of molecular connectivity index, C_{ij} for each bond is calculated using formula $C_{ij}=\left(\delta_i\delta_j\right)_k^{-1/2}$ such that summation of all these C_{ij} values for entire molecule gives molecular connectivity index.

As stated above, the molecular connectivity index does not specify bonding type between adjacent atoms. However, in the valence connectivity index developed by Kier and Hall, 1981, the number of sigma, pi and lone pair orbitals (again except bonds with hydrogen) are considered. The adjacency number, in this case, is expressed as δ^v and resulting connectivity index would be as Equation 5.12.

$$^1X^v=\sum\left(\delta_i^v\delta_j^v\right)^{-1/2} \tag{4.12}$$

The least significant descriptor in the model, GATS2v, which is a 2D autocorrelations descriptor, positively correlated with the endpoint. It is an autocorrelation descriptor weighted by van der Waals volume such that it can be claimed that the volume of a molecule has an effect on its cytotoxicity measured with NR assay.

Geary autocorrelation descriptors have been used in previous studies for modeling of antifungal activity (Duchowicz et al., 2007), modeling of certain enzymatic activities (Amin et al., 2016) and

for QSTR modeling of toxicity of esters to *Daphnia magna* by Papa and co-workers (2005). Moreover, Tuğcu and co-workers (2012) have also found a positive relationship between another 2D autocorrelations descriptor, GATS3p, and toxicity of pharmaceuticals to fish.

5.4.4. Correlation between $pEC_{50,NR[FHM]}$ and log K_{ow}

The log K_{ow} values ranged from -1.68 for citric acid to 2.64 for *n*-octanol in the $pEC_{50,NR[FHM]}$ dataset. When the regression analysis carried out, it was shown that there is no statistically significant correlation between log K_{ow} and $pEC_{50,NR[FHM]}$ for this dataset (n = 35; r = 0.166; p = 0.339). The plot of observed $pEC_{50,NR[FHM]}$ values versus log K_{ow} values visualizing the relationship is given in Figure 4.20.

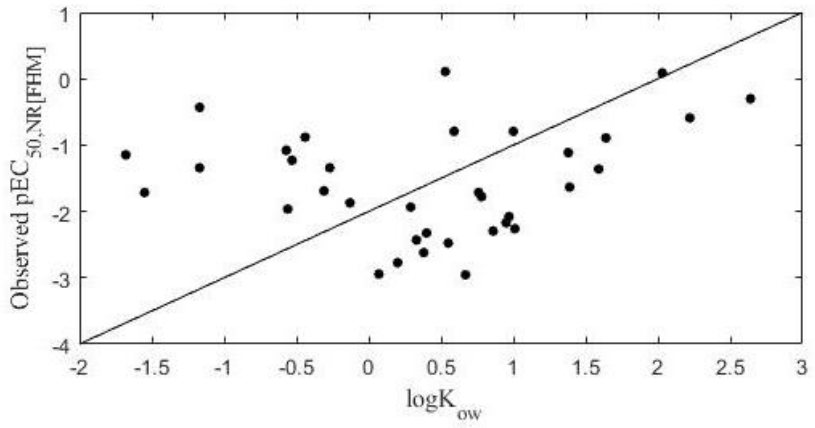


Figure 4.20. Plot of observed $pEC_{50,NR[FHM]}$ values versus log K_{ow} values.

4.4.5. Correlation of Experimental *in vivo* pLC_{50} with Predicted *in vitro* $pEC_{50,NR[FHM]}$

Correlation analysis was carried out to evaluate the relationship between the predicted $pEC_{50,NR[FHM]}$ values and the fish lethal toxicity (pLC_{50}) data compiled from the literature were evaluated. As a result of correlation analysis, no statistically significant correlation between *in vitro* $pEC_{50,\ NR[FHM]}$ and experimental *in vivo* pLC_{50} values was found. As stated before, the dataset noncytotoxic nature of the data set may lead to this situation. It can be claimed that the model predictions would be only valid for nontoxic external set chemicals yet pLC_{50} data covered a broad range of toxicity values.

4.5. QSTR Modelling for $pEC_{50,MTT[PLHC-1]}$

4.5.1. Dataset

Experimental $pEC_{50,MTT[PLHC-1]}$ values and related physicochemical properties of chemicals used for $pEC_{50,MTT[PLHC-1]}$ modelling are listed in Table 4.32.

Table 4.32. The name, CAS number, log K_{ow}, molecular weight and experimental $pEC_{50,MTT[PLHC-1]}$ values of chemicals in the dataset.

ID	Compound Name	CAS number	log K_{ow}*	Molecular Weight**	$pEC_{50,MTT[PLHC-1]}$ (mM)	References
1	*(S)*-propranolol	4199-09-1	0.36	259.349	0.82	Caminada et al., 2006
2	2,3,4,5-tetrachlorophenol	4901-51-3	3.88	231.893	0.59	Fent and Hunn, 1996
3	2,3,6-trichlorophenol	933-75-5	3.32	197.448	0.68	Fent and Hunn, 1996
4	2,4,5-trichlorophenol	95-95-4	3.32	197.448	0.00	Fent and Hunn, 1996
5	2,4-dichlorophenol	120-83-2	2.76	163.003	0.57	Fent and Hunn, 1996
6	2,4-dinitrophenol	51-28-5	1.71	184.107	0.11	Fent and Hunn, 1996
7	2,5-dinitrophenol	329-71-5	1.71	184.107	0.20	Fent and Hunn, 1996
8	2,6-dichlorophenol	87-65-0	2.76	163.003	-0.74	Fent and Hunn, 1996
9	2-chlorophenol	95-57-8	2.20	128.558	-0.11	Fent and Hunn, 1996
10	2-methyl-4,6-dinitrophenol	534-52-1	2.20	198.134	0.21	Fent and Hunn, 1996
11	2-nitrophenol	88-75-5	1.68	139.110	-1.08	Fent and Hunn, 1996
12	2-*s*-butyl-4,6-dinitrophenol	88-85-7	3.36	240.215	0.66	Fent and Hunn, 1996
13	2-*t*-butyl-4,6-dinitrophenol	1420-07-1	3.42	240.215	0.66	Fent and Hunn, 1996
14	3,4-dichlorophenol	95-77-2	2.76	163.003	0.14	Fent and Hunn, 1996
15	3-nitrobenzene sulfonic acid	98-47-5	1.23	203.174	-2.20	Fent and Hunn, 1996
16	3-nitrophenol	554-84-7	1.68	139.110	-0.34	Fent and Hunn, 1996
17	4-aminonaphthalene - 1-sulfonic acid	84-86-6	1.39	223.252	-2.48	Fent and Hunn, 1996
18	4-chloro-2-methylphenol	1570-64-5	2.69	142.585	-0.04	Fent and Hunn, 1996
19	4-chlorophenol	106-48-9	2.20	128.558	-0.38	Fent and Hunn, 1996
20	4-nitrophenol	100-02-7	1.68	139.110	-0.60	Fent and Hunn, 1996
21	4-nonylphenol	104-40-5	5.47	220.356	0.16	Fent and Hunn, 1996
22	4-octylphenol	1806-26-4	5.05	206.329	0.17	Fent and Hunn, 1996
23	Atorvastatin	134523-00-5	3.41	558.650	1.09	Caminada et al., 2006
24	Bezafibrate	41859-67-0	3.77	361.825	-0.41	Caminada et al., 2006
25	Clofibric acid	882-09-7	2.62	214.648	-0.58	Caminada et al., 2006
26	Diazepam	439-14-5	3.27	298.773	0.44	Caminada et al., 2006
27	Diclofenac	15307-86-5	3.60	296.153	0.61	Caminada et al., 2006
28	Doxorubicin	23214-92-8	-0.29	527.526	2.59	Caminada et al., 2006
29	Fluoxetine	54910-89-3	4.44	309.331	1.69	Caminada et al., 2006
30	Furosemide	54-31-9	-0.14	330.748	-0.53	Caminada et al., 2006
32	Hydrochlorothiazide	58-93-5	-3.89	297.743	-0.30	Caminada et al., 2006
33	Ibuprofen	15687-27-1	3.75	206.285	-0.08	Caminada et al., 2006

Table 4.32. Continued.

ID	Compound Name	CAS number	Log K_{ow}*	Molecular Weight**	$pEC_{50,MTT[PLHC-1]}$ (mM)	References
34	Mefenamic acid	61-68-7	3.52	241.290	0.41	Caminada et al., 2006
35	Naproxen	22204-53-1	2.97	230.263	-0.40	Caminada et al., 2006
36	OH-tamoxifen	68047-06-3	5.67	387.523	1.86	Caminada et al., 2006
37	Pentachlorophenol	87-86-5	4.44	266.338	0.49	Fent and Hunn, 1996
38	Phenol	108-95-2	1.64	94.113	-0.78	Fent and Hunn, 1996
39	Propranolol	525-66-6	0.36	259.349	0.80	Caminada et al., 2006
40	Rofecoxib	162011-90-7	2.67	316.377	0.27	Caminada et al., 2006
41	Simvastatin	79902-63-9	4.56	418.574	1.19	Caminada et al., 2006
42	Tamoxifen	10540-29-1	6.05	371.524	1.70	Caminada et al., 2006

**n*-octanol water partition coefficients calculated by SPARTAN 10; ** MW calculated by SPARTAN 10

4.5.2. Model Development

Best 3-descriptor QSTR models were selected from five randomly divided sets. Test set chemicals in each of five divisions with the ratio of the number of test set to training set compounds were given in Table 4.33. Additionally, fit, internal and external validation parameters for the selected models are given in Table 4.34 and 4.35, respectively. In Tables 4.34 and 4.35, 1-2 descriptor combinations that were used to build the selected models on were included, as well. In this way, how the addition of each new descriptor contributed to the model can be seen.

Table 4.33. Test set compounds for five divisions used in the QSTR modelling of $pEC_{50,MTT[PLHC-1]}$.

Division no	n_{TEST}/n_{TR}	Test Set Compounds*
1	10/32	2, 3, 10, 18, 21, 24, 29, 30, 33, 38
2	12/30	2, 3, 13, 14, 21, 24, 29, 30, 32, 36, 38, 41
3	10/32	2, 4, 7, 12, 16, 20, 25, 26, 29, 42
4	10/32	11, 14, 19, 20, 22, 25, 29, 36, 37, 41
5	10/32	2, 4, 6, 19, 20, 26, 29, 38, 39, 41

*Compound numbers refer to the ID numbers given in Table 4.32.

Table 4.34. Fit and internal validation parameters of the generated $pEC_{50,MTT[PLHC-1]}$ models.

No	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
				Division 1					
1.1	VE1_H2 SpMax1_Bh(e) CATS2D_01_AN	0.826	0.807	0.419	0.905	44.222	0.775	0.476	0.875
1.2	VE1_H2 SpMax1_Bh(e) CATS2D_02_NL	0.784	0.761	0.467	0.879	33.900	0.693	0.556	0.827
1.3	VE1_H2 CATS2D_01_AN	0.749	0.732	0.503	0.857	43.336	0.683	0.565	0.814
1.4	VE1_H2 CATS2D_02_NL	0.623	0.597	0.616	0.768	23.981	0.470	0.731	0.681
1.5	VE1_H2	0.372	0.351	0.796	0.542	17.735	0.298	0.841	0.482
				Division 2					
2.1	VE1_H2 SpMax1_Bh(e) CATS2D_02_NL	0.769	0.742	0.462	0.869	28.842	0.669	0.553	0.809
				Division 3					
***3.1**	**Mor28e NaasC CATS2D_01_AN**	**0.775**	**0.751**	**0.460**	**0.874**	**32.222**	**0.702**	**0.529**	**0.831**
3.2	Mor28e NaasC CATS2D_01_DN	0.708	0.676	0.525	0.829	22.574	0.601	0.613	0.768
3.3	Mor28e NaasC	0.498	0.464	0.687	0.665	14.401	0.357	0.778	0.577
3.4	Mor28e CATS2D_01_AN	0.459	0.422	0.714	0.629	12.301	0.342	0.787	0.535
3.5	Mor28e CATS2D_01_DN	0.338	0.293	0.789	0.505	7.409	0.193	0.872	0.400
3.6	Mor28e	0.222	0.196	0.856	0.363	8.543	0.125	0.908	0.283
				Division 4					
4.1	VE1_H2 Eig01_AEA(dm) CATS2D_01_AN	0.847	0.830	0.363	0.917	51.481	0.796	0.418	0.888
4.2	VE1_H2 SpMax_AEA(dm) CATS2D_01_AN	0.847	0.830	0.363	0.917	51.481	0.796	0.418	0.888
4.3	VE1_H2 CATS2D_01_AN	0.722	0.702	0.489	0.838	37.556	0.645	0.552	0.786
4.4	VE1_H2	0.235	0.210	0.810	0.381	9.235	0.121	0.868	0.272

Table 4.34. (Continued).

No	Descriptors	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
	Division 5								
5.1	VE1_A JGI5 CATS2D_01_AN	0.832	0.814	0.403	0.908	46.298	0.774	0.468	0.875
5.2	VE1_A CATS2D_01_AN	0.746	0.728	0.496	0.854	42.520	0.687	0.551	0.816
5.3	VE1_A	0.246	0.221	0.854	0.395	9.776	0.139	0.913	0.299

*The selected model.

Table 4.35. External Validation Parameters of the generated $pEC_{50,MTT[PLHC-1]}$ models.

No	**Descriptors**	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2-r_0'^2)/r^2$	$(r^2-r_0^2)/r^2$
	Division 1												
1.1	VE1_H2 SpMax1_Bh(e) CATS2D_01_AN	0.811	0.307	0.794	0.794	0.906	0.881	0.732	0.028	0.751	1.075	0.008	0.016
1.2	VE1_H2 SpMax1_Bh(e) CATS2D_02_NL	0.788	0.320	0.776	0.776	0.898	0.877	0.716	0.031	0.784	1.003	0.007	0.016
1.3	VE1_H2 CATS2D_01_AN	0.721	0.427	0.603	0.603	0.820	0.766	0.480	0.011	0.597	1.043	0.149	0.163
1.4	VE1_H2 CATS2D_02_NL	0.567	0.524	0.401	0.400	0.727	0.661	0.344	0.004	0.519	0.850	0.278	0.269
1.5	VE1_H2	0.155	0.687	-0.027	-0.029	0.533	0.346	0.118	0.009	0.242	0.520	0.280	0.463
	Division 2												
2.1	VE1_H2 SpMax1_Bh(e) CATS2D_02_NL	0.804	0.366	0.831	0.802	0.855	0.893	0.753	0.080	0.875	0.963	0.016	0.000
	Division 3												
***3.1**	**Mor28e NaasC CATS2D_01_AN**	**0.864**	**0.290**	**0.878**	**0.864**	**0.911**	**0.926**	**0.829**	**0.055**	0.887	1.003	**0.006**	**0.000**
3.2	Mor28e NaasC CATS2D_01_DN	0.861	0.303	0.867	0.851	0.903	0.928	0.834	0.052	0.953	0.928	0.000	0.005
3.3	Mor28e NaasC	0.833	0.400	0.768	0.740	0.830	0.853	0.642	0.063	0.688	1.172	0.044	0.086
3.4	Mor28e CATS2D_01_AN	0.686	0.459	0.695	0.658	0.776	0.764	0.587	0.124	0.620	1.197	0.080	0.004

Table 4.35. Continued.

No	Descriptors	R^2_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2m av.	Δr^2m	k'	k	$(r^2-r_0'^2)/r^2$	$(r^2-r_0^2)/r^2$
						Division 3							
3.5	Mor28e CATS2D_01_DN	0.682	0.462	0.690	0.653	0.773	0.771	0.629	0.028	0.615	1.201	0.014	0.005
3.6	Mor28e	0.545	0.589	0.497	0.437	0.631	0.629	0.443	0.105	0.452	1.248	0.015	0.148
						Division 4							
4.1	VE1_H2 Eig01_AEA(dm) CATS2D_01_AN	0.757	0.474	0.762	0.755	0.738	0.855	0.711	0.064	0.749	1.037	0.014	0.000
4.2	VE1_H2 SpMax_AEA(dm) CATS2D_01_AN	0.757	0.474	0.762	0.755	0.738	0.855	0.711	0.064	0.749	1.037	0.014	0.000
4.3	VE1_H2 CATS2D_01_AN	0.774	0.466	0.770	0.764	0.747	0.853	0.694	0.066	0.720	1.096	0.028	0.005
4.4	VE1_H2	0.735	0.644	0.561	0.549	0.517	0.661	0.587	0.045	0.415	1.707	0.040	0.073
						Division 5							
5.1	VE1_A JGI5 CATS2D_01_AN	0.745	0.410	0.718	0.702	0.826	0.857	0.680	0.087	0.948	0.823	0.029	0.001
5.2	VE1_A CATS2D_01_AN	0.726	0.409	0.719	0.703	0.827	0.840	0.612	0.104	0.859	0.884	0.072	0.010
5.3	VE1_A	0.726	0.528	0.533	0.506	0.712	0.701	0.435	0.060	0.486	1.228	0.178	0.270

*The selected model.

4.5.3. Comparison of Applicability Domain of $pEC_{50,MTT[PLHC-1]}$ Models

The best QSTR model for $pEC_{50,MTT[PLHC-1]}$ was selected by comparing the structural coverage of the generated three-descriptor QSTR models. With this purpose, an external set of 660 chemicals, without experimental $pEC_{50,MTT[PLHC-1]}$ values, were used. Number of chemicals that the model could predict $pEC_{50,MTT[PLHC-1]}$ values out of 660 were given in Table 4.36 together with the corresponding structural coverage (%).

Table 4.36. Predictive performances of the generated $pEC_{50,MTT[PLHC-1]}$ models.

Division no	Model no	Number of compounds within the AD of $pEC_{50,MTT[PLHC-1]}$ model (out of 660)	Structural Coverage (%)
1	1.1	449	68.0
	1.2	440	66.7
2	2.1	433	65.6
3	*3.1	**649**	**98.3**
	3.2	636	96.4
4	4.1	581	88.0
	4.2	581	88.0
5	5.1	612	92.7

*The selected model.

Depending on structural coverage, model 3.1 was found to be superior to other models therefore, it was selected as the best model. Fit, internal validation parameters, external validation parameters and structural coverage of model 3.1 were written in bold through Tables 4.34 - 4.36, respectively. The mathematical QSTR equation for the model 3.1 is given in Eq. 4.13 where numbers in parentheses show the 95% confidence interval for variable coefficients in the equation.

$$pEC_{50,MTT[PLHC-1]} = -1.370\ (\pm 0.506)\ Mor28e + 0.301\ (\pm 0.098)\ NaasC + 0.858\ (\pm 0.299)\ CATS2D\_01\_AN - 0.780\ (\pm 0.397) \quad (4.13)$$

The selected model had satisfying performance regarding all the fit, internal and external criteria. The fit of the model was demonstrated with high $R^2 = 0.775$, $R^2_{adj} = 0.751$ and low $RMSE_{TR} = 0.460$ values. Its robustness and stability were verified with a satisfying value of Q^2_{LOO} (0.874). Significantly low R^2_{Yscr} and Q^2_{Yscr} values (0.095 and -0.193, respectively) eliminated the possibility that the model was obtained by chance correlation. High R^2_{TEST} =0.864 and low $RMSE_{TEST} = 0.290$ values testified the predictive ability of the model. Additionally, the model performed higher values than the corresponding literature threshold values for the external validation parameters Q^2_{F1}, $Q^2_{F2,}$ Q^2_{F3} and CCC_{TEST} by which the predictive power of the model justified. Regarding all the other external validation parameters specified in Table 4.35, the model demonstrated satisfying results.

Furthermore, when the Golbraikh and Tropsha's criteria were employed, it was investigated that the model was able to fulfill all the expected conditions. Finally, the predictive ability of the model was also approved as good by MAE-based criteria (Roy et al., 2016) with values of *MAE* (95% of the data) = 0.265, $3\sigma = 0.373$ and training set range = 4.291.

Chemicals used for the QSTR modelling of $pEC_{50,MTT[PLHC-1]}$ endpoint together with their training/test set status in modelling, experimental and predicted $pEC_{50,MTT[PLHC-1]}$ values, hat and descriptor values are given in Table 4.37.

Table 4.37. Chemicals used for the QSTR modelling of $pEC_{50,MTT[PLHC-1]}$, their experimental and predicted $pEC_{50,MTT[PLHC-1]}$ values, hat values and descriptor values.

Chemicals	Status	Exp. $pEC_{50,MTT[PLHC-1]}$ (mM)	Pred. $pEC_{50,MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h*=0.375)	Mor28e	NaasC	CATS2D_01_AN
(S)-propranolol	Training	0.820	0.035	0.091	-0.375	1	0
2,3,4,5-tetrachlorophenol	Test	0.590	0.391	0.112	0.244	5	0
2,3,6-trichlorophenol	Training	0.680	0.264	0.070	0.117	4	0
2,4,5-trichlorophenol	Test	0.000	0.201	0.077	0.163	4	0
2,4-dichlorophenol	Training	0.570	0.034	0.060	0.065	3	0
2,4-dinitrophenol	Training	0.110	-0.082	0.073	0.150	3	0
2,5-dinotrophenol	Test	0.200	-0.110	0.077	0.170	3	0
2,6-dichlorophenol	Training	-0.740	0.045	0.059	0.057	3	0
2-chlorophenol	Training	-0.110	-0.111	0.063	-0.049	2	0
2-methyl-4,6-dinitrophenol	Training	0.210	0.222	0.075	0.148	4	0
2-nitrophenol	Training	-1.080	-0.250	0.075	0.053	2	0
2-*s*-butyl-4,6-dinitrophenol	Test	0.660	0.587	0.051	-0.119	4	0
2-*t*-butyl-4,6-dinitrophenol	Training	0.660	0.898	0.059	-0.346	4	0
3,4-dichlorophenol	Training	0.140	0.078	0.057	0.033	3	0
3-nitrobenzene sulfonic acid	Training	-2.200	-2.001	0.297	0.078	2	2
3-nitrophenol	Test	-0.340	-0.064	0.061	-0.083	2	0
4-aminonaphthalene-1-sulfonic acid	Training	-2.480	-2.163	0.312	0.196	2	2
4-chloro-2-methylphenol	Training	-0.040	0.164	0.050	-0.030	3	0
4-chlorophenol	Training	-0.380	-0.105	0.063	-0.053	2	0
4-nitrophenol	Test	-0.600	-0.141	0.065	-0.027	2	0
4-nonylphenol	Training	0.160	0.569	0.083	-0.545	2	0
4-octylphenol	Training	0.170	0.426	0.068	-0.441	2	0

Table 4.37. Continued.

Chemicals	Status	Exp. $pEC_{50,MTT[PLHC-1]}$ (mM)	Pred. $pEC_{50,MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h*=0.375)	Mor28e	NaasC	CATS2D_01_AN
Atorvastatin	Training	1.090	1.454	0.323	-0.499	8	1
Bezafibrate	Training	-0.410	0.447	0.139	-0.643	4	1
Clofibric acid	Test	-0.580	-0.851	0.084	-0.135	2	1
Diazepam	Test	0.440	0.730	0.051	-0.223	4	0
Diclofenac	Training	0.610	-0.344	0.104	0.154	5	1
Doxorubicin	Training	2.590	2.045	0.278	-0.304	8	0
Fluoxetine	Test	1.690	1.244	0.143	-0.818	3	0
Furosemide	Training	-0.530	-0.666	0.146	0.389	5	1
Gemfibrozil	Training	0.060	-0.123	0.090	-0.447	3	1
Hydrochlorothiazide	Training	-0.300	-0.217	0.152	0.468	4	0
Ibuprofen	Training	-0.080	-0.581	0.091	-0.332	2	1
Mefenamic acid	Training	0.410	0.110	0.091	-0.177	5	1
Naproxen	Training	-0.400	-0.581	0.091	-0.332	2	1
OH-tamoxifen	Training	1.860	2.128	0.274	-1.024	5	0
Pentachlorophenol	Training	0.490	0.542	0.174	0.354	6	0
Phenol	Training	-0.780	-0.289	0.090	-0.139	1	0
Propranolol	Training	0.800	0.255	0.108	-0.536	1	0
Rofecoxib	Training	0.270	0.678	0.055	-0.405	3	0
Simvastatin	Training	1.190	0.476	0.241	-0.917	0	0
Tamoxifen	Test	1.700	1.822	0.235	-1.020	4	0

Figure 4.21 shows predicted $pEC_{50,MTT[PLHC-1]}$ values from Eq 4.13 versus experimental $pEC_{50,MTT[PLHC-1]}$ values. As data points were scattered well along the optimal line, it can be claimed that the experimental and predicted $pEC_{50,MTT[PLHC-1]}$ values are in good agreement.

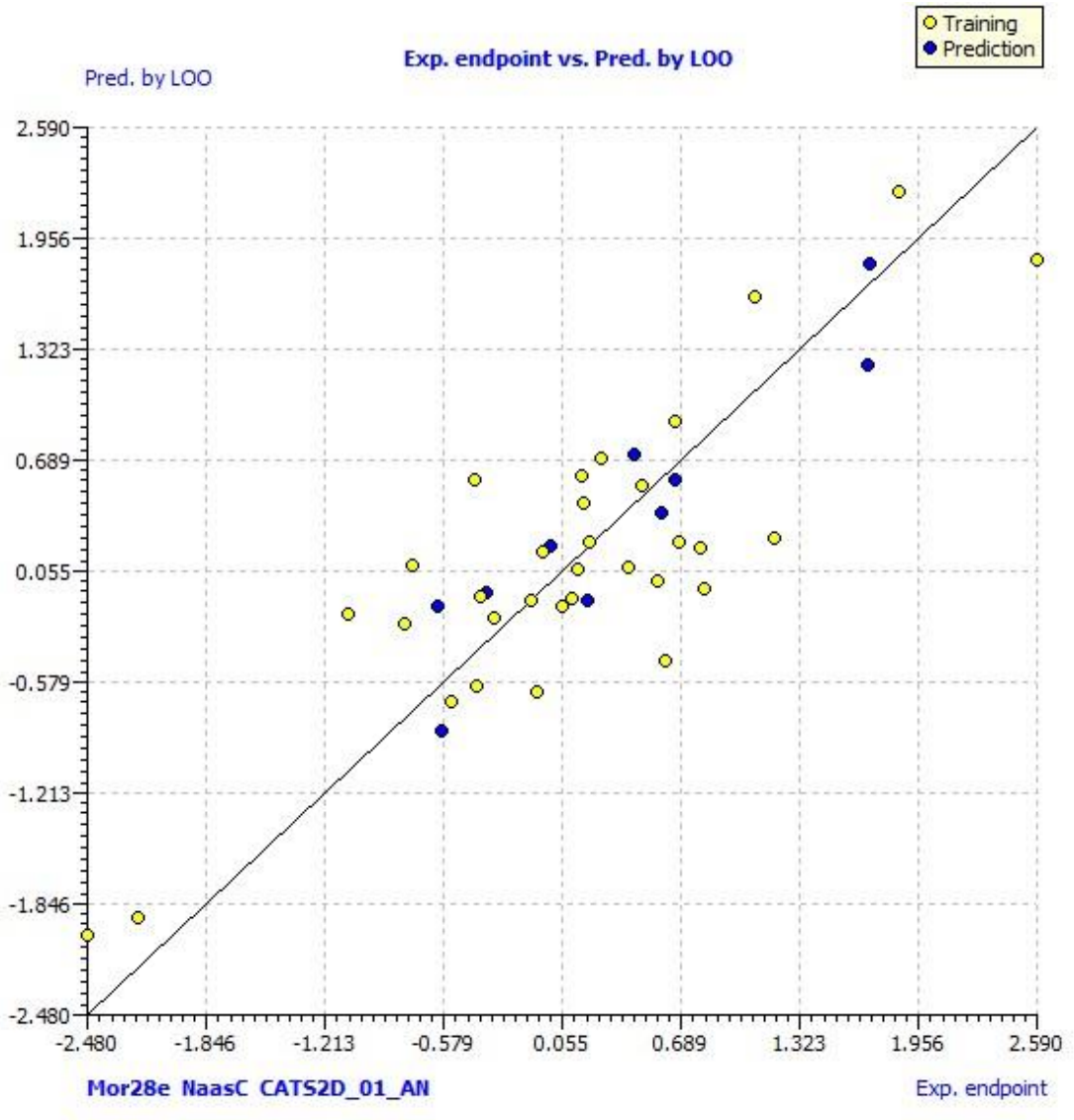


Figure 4.21. Plot of predicted $pEC_{50,MTT[PLHC-1]}$ values from Eq. 4.13 versus experimental $pEC_{50,MTT[PLHC-1]}$ values.

Williams plot indicating AD of $pEC_{50,MTT[PLHC-1]}$ model is given in Figure 4.22. It is obvious that all the training and test set compounds had residuals lower than ±2.5σ and none of them had hat values greater than the critical hat value (h^* = 0.375) demonstrating the presence of neither response nor structural outliers, respectively.

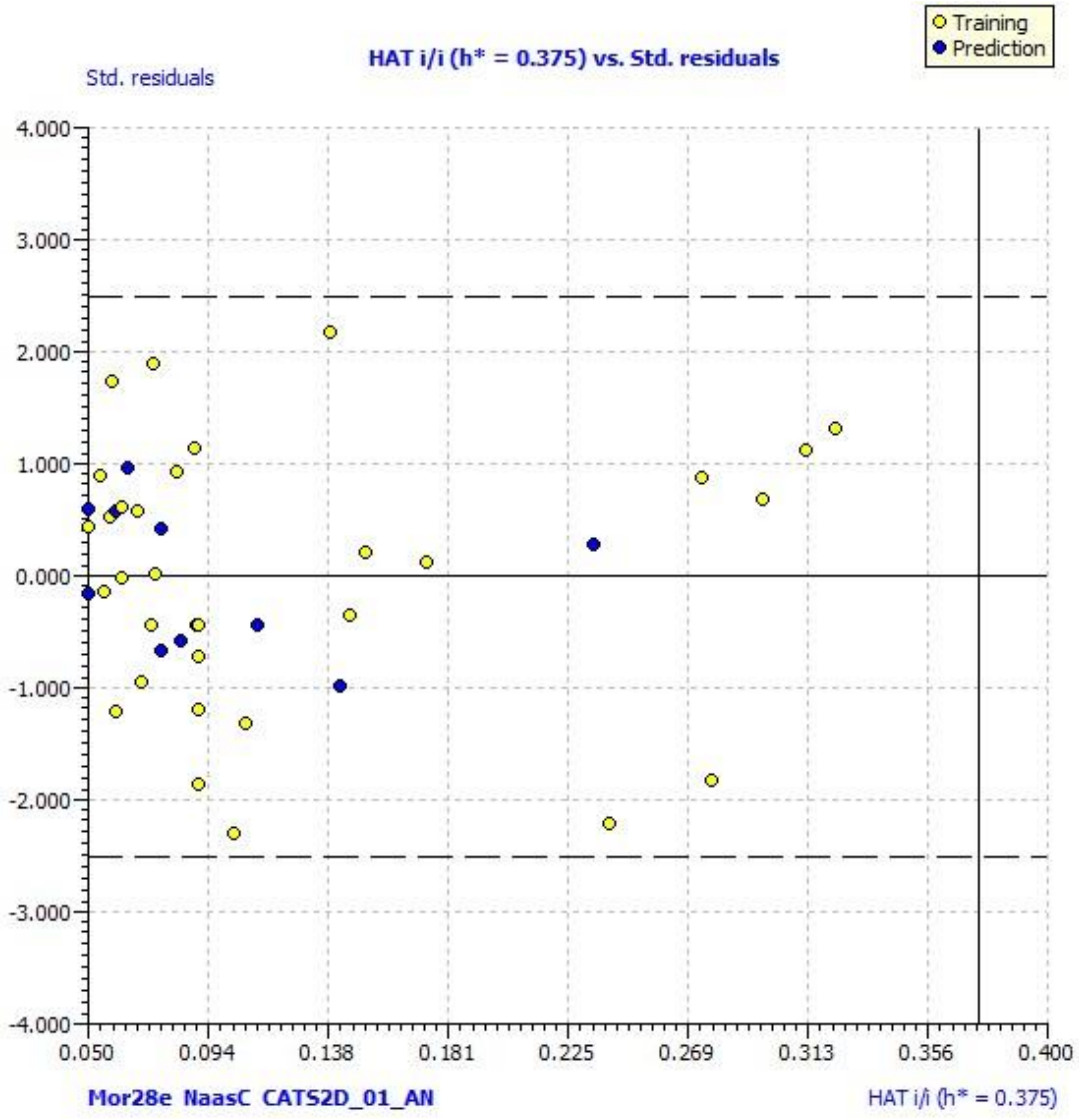


Figure 4.22. Williams plot for $pEC_{50,MTT[PLHC-1]}$ model.

To test the external predictive ability of the $pEC_{50,MTT[PLHC-1]}$ model, an external set of 660 chemicals was used. These chemicals with their calculated descriptor, predicted and experimental $pEC_{50,MTT[PLHC-1]}$ values are given in Appendix E. The plot of predicted $pEC_{50,MTT[PLHC-1]}$ values for training, test and external set chemicals versus their hat values (Insubria graph) is provided in Figure 4.23.

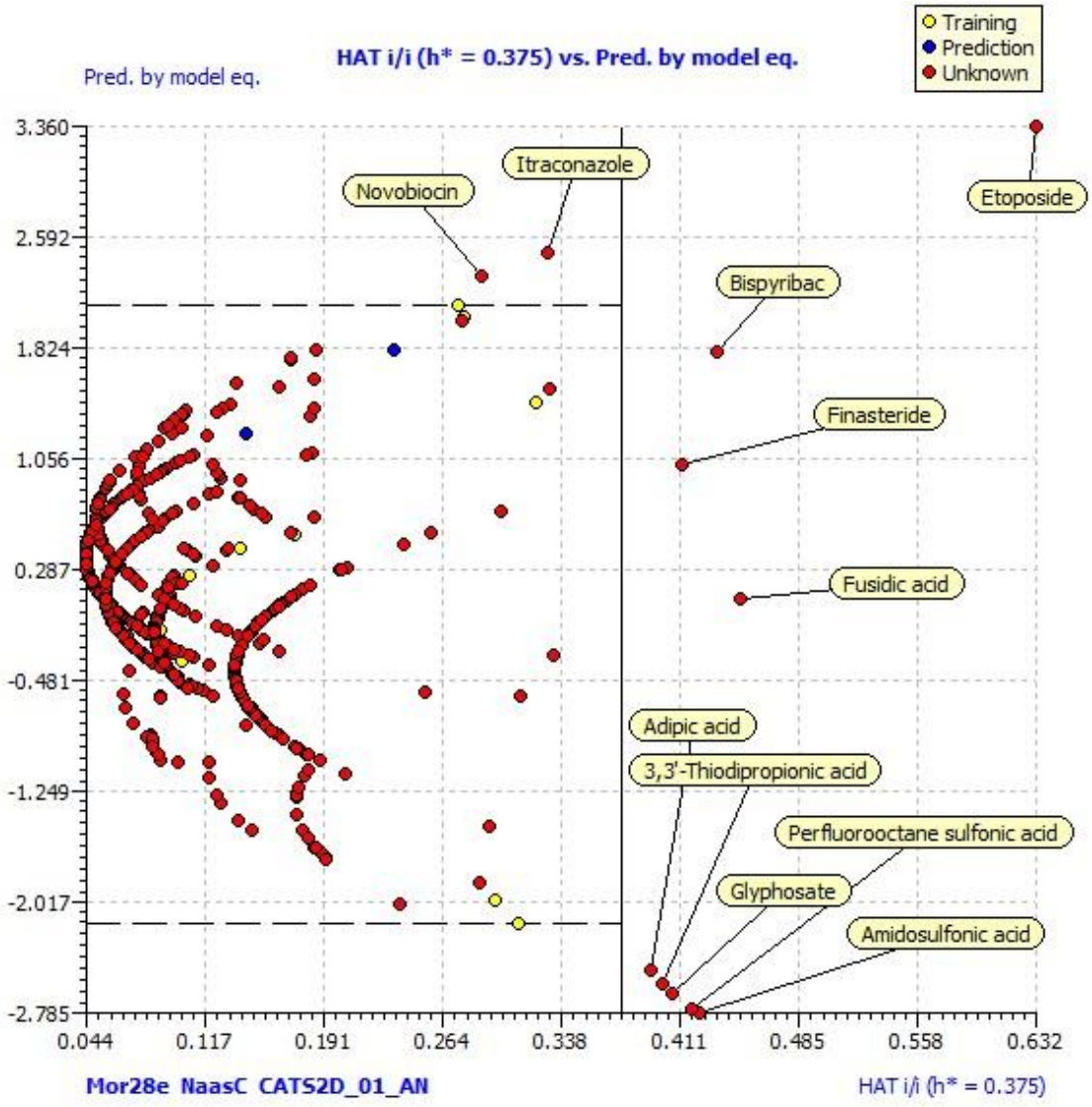


Figure 4.23. Insubria graph of $pEC_{50,MTT[PLHC-1]}$ model. Predicted $pEC_{50,MTT[PLHC-1]}$ values for training, test and external (660 chemicals) set chemicals from Eq. 4.13 versus their hat values.

The most and the least toxic ones among the external set chemicals were screened based on predicted $pEC_{50,MTT[PLHC-1]}$ values (Table 4.38). Depending on predicted $pEC_{50,MTT[PLHC-1]}$ values, tetrabromobisphenol A was found as the most cytotoxic chemical. Considering European Union Risk Assessment Report indicating tetrabromobisphenol A as a very toxic chemical depending on its long-term adverse effects on aquatic organisms (ECHA, 2006).

Table 4.38. The most and the least cytotoxic chemicals from the external set with predicted by Eq. 4.13.

	Name	Chemical class	$pEC_{50,\ MTT[PLHC\text{-}1]}$ pred. by Eq. 4.13 (mM)
Most Toxic	Tetrabromobisphenol A	Diphenyl alkane, dihydroxy,bromo	2.023
	Glibenclamide	Pharmaceutical, Antidiabetic Agents	1.823
	Pyrazosulfuron ethyl	Pesticide, Herbicide	1.758
	Prazosin	Pharmaceutical, Antihypertensive Agents	1.756
	Trimethoprim	Phenyl-diazine,diamino	1.621
	Cyclosulfamuron	Pesticide, Herbicide	1.594
	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	Diphenyl alkane,alkoxy-alcohol	1.565
	Quinidine	Pharmaceutical, Antiarrhythmic Agents	1.552
	Terazosin	Pharmaceutical, Antihypertensive Agents	1.437
	2-(1,1-dimethyl)-4,6-dimethylphenol	Phenol	1.411
Least Toxic	Perfluorooctanoic acid	Acid, polyfluoro	-2.024
	4-nitrotoluene-2-sulphonic acid	Benzene, sulfonic acid, nitro	-1.874
	Glyoxylic acid	Acid-aldehyde	-1.709
	A-fluoro-b-alanine	Acid, amino, fluoro	-1.698
	Chloroacetic acid	Acid, chloro	-1.678
	Sorbic acid	Acid, a,b-unsaturated	-1.640
	Acrylic acid	Acid, a,b-unsaturated	-1.629
	Methacrylic acid	Acid, a,b-unsaturated	-1.570
	Cefuroxime	Pharmaceutical, Antiinfective Agents	-1.518
	2-methyl butanoic acid	Acid	-1.518

Tetrabromobisphenol A is frequently used reactive and/or additive flame retardant and about 40,000 tons of tetrabromobisphenol A is imported to EU per year (ECHA, 2006). Therefore, it would be beneficial to further assess its ecotoxic potencies and regulate its usage/disposal. Furthermore, in accordance with the $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,NR[GFS]}$ predictions, most toxic chemicals were from pharmaceuticals (4 of 10) thus the prioritization of this type of chemicals could be further highlighted.

Molecular descriptors appeared the $pEC_{50,MTT[PLHC-1]}$ model and their standardized coefficients, indicating the relative importance of each descriptor are given in Table 4.39.

Table 4.39. Descriptors appeared in the $pEC_{50,MTT[PLHC-1]}$ model.

Descriptor	Meaning of descriptor	Type	Standard coefficient
NaasC	Number of atoms of type aasC Atom-type	E-state indices	0.570
CATS2D_01_AN	CATS2D Acceptor-Negative at lag 01	CATS 2D	-0.530
Mor28e	signal 28 / weighted by Sanderson electronegativity	3D-MoRSE descriptors	-0.503

The most important descriptor in the model, NaasC positively contributed to the $pEC_{50,MTT[PLHC-1]}$ endpoint. The same descriptor also appeared in the $pEC_{50,NR[PLHC-1]}$ model reported in the present study. In this way, the importance of the information encoded by the descriptor in explaining the cytotoxicity can be further justified.

On the other hand, CATS2D_01_AN is a 2D structure-based atom-pair descriptor encoding topological information where CATS denotes a chemically advanced template search and AN refers to an hydrogen-bond acceptor, either negatively charged or ionizable (Todeschini and Consonni, 2009). As it is based on hydrogen-bond acceptors in a molecule, it encodes an information relevant to electronegativity. Therefore, considering the positive contribution of CATS2D_01_AN in the $pEC_{50,MTT[PLHC-1]}$ model, it can be deduced that as the electronegativity of a molecule increases, metabolic impairment increases.

The least important descriptor, Mor28e, belongs to the group 3D-MoRSE descriptors. 3D-MoRSE (Molecule Representation of Structures based on Electron diffraction) descriptors are based on the three-dimensional structure of molecules by a certain number of values. These descriptors are

derived from 3D-atomic coordinates which are transformed into molecular codes by mathematical transformations used in electron diffraction (Schuur et al., 1996). The mathematical transformation developed by Schuur and co-workers (1996) is given in Eq. 4.14.

$$I(s)=\sum_{i=2}^{N}\sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}} \quad (4.14)$$

In this transformation, equation s represents scattering angle which has values between 0 and 31.0 $Å^{-1}$. These 32 values constitute the 3D-MoRSE code of a molecule. A_i and A_j represent properties of atoms at i and j positions and r_{ij} represents interatomic distances.

In a modelling study of pharmaceuticals to fish, Tugcu and co-workers (2012) have used 3D-MoRSE descriptors and found out that 3D-MoRSE descriptors are related to fish toxicity because of its correlation with hydrophobicity. In addition, Caballero and Fernandez (2006) have also used a 3D-MoRSE descriptor for antifungal activity modeling.

The model endpoint, $pEC_{50,MTT[PLHC-1]}$ is a measure of metabolic impairment to a target cell or in other words, a measure of the metabolic capability of fish cells. As metabolic processes require interatomic relationships and electrochemical reactions, a descriptor dealing with electronic and atomic properties of a molecule would be interpretable for modeling of such endpoint. Additionally, the presence of a descriptor related with electronic and atomic properties positively correlating to the target endpoint coincides with the previous findings of Önlü and Saçan (2017) in which it was stated that toxicity of molecules tends to have a positive correlation with their electrophilic character.

4.5.4. Correlation between $pEC_{50,MTT[PLHC-1]}$ and log K_{ow}

The log K_{ow} values ranged from -3.89 for hydrochlorothiazide to 6.05 for tamoxifen. Depending on correlation analysis a weak correlation was found between log K_{ow} and $pEC_{50,MT[PLHC-1]}$ for the data set chemicals (n = 42; r = 0.320; p < 0.05). This is probably due to chemicals with complex structures and different mode of action appeared in the data set. The relationship between observed $pEC_{50,MT[PLHC-1]}$ values and log K_{ow} is given in Figure 4.24.

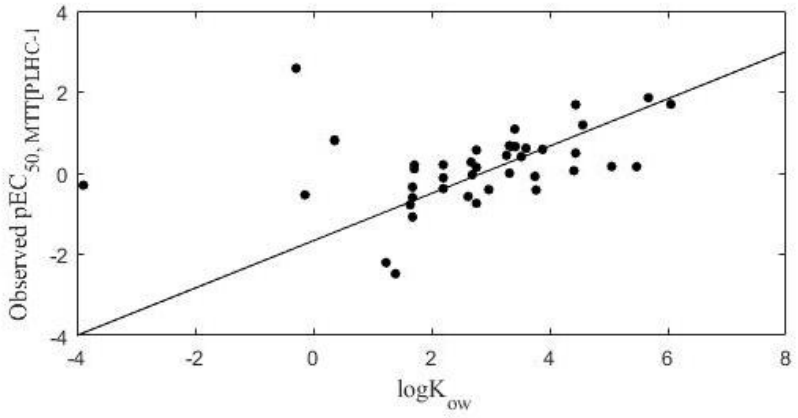


Figure 4.24. Plot of observed $pEC_{50,MTT[PLHC-1]}$ values versus $logK_{ow}$ values.

4.5.5. Correlation of Experimental *in vivo* pLC_{50} with Predicted *in vitro* $pEC_{50,MTT[PLHC-1]}$

In order to conclude the usability of the **$pEC_{50,MTT[PLHC-1]}$** model as an alternative to *in vivo* toxicity testing, the relationship between the predicted $pEC_{50MTT\ [PLHC-1]}$ values by Eq. 4.13 and the fish lethality (pLC_{50}) data compiled from the literature was assessed. The experimental pLC_{50} and predicted $pEC_{50,MTT[PLHC-1]}$ values for external set chemicals are given in Appendix E.

A moderate correlation (n = 184; r = 0.500; $p < 0.01$) was found between the predicted *in vitro* $pEC_{50,MTT[PLHC-1]}$ and the experimental *in vivo* pLC_{50} data. Plot of observed pLC_{50} versus predicted $pEC_{50,MTT[PLHC-1]}$ values was given in Figure 4.25. Among the 204 chemicals, 20 chemicals (mainly from bromo- and chloro- substituted alkanes similar to the case of $pEC_{50,NR[PLHC-1]}$ predictions) showed a weak relationship with the experimental pLC_{50} values.

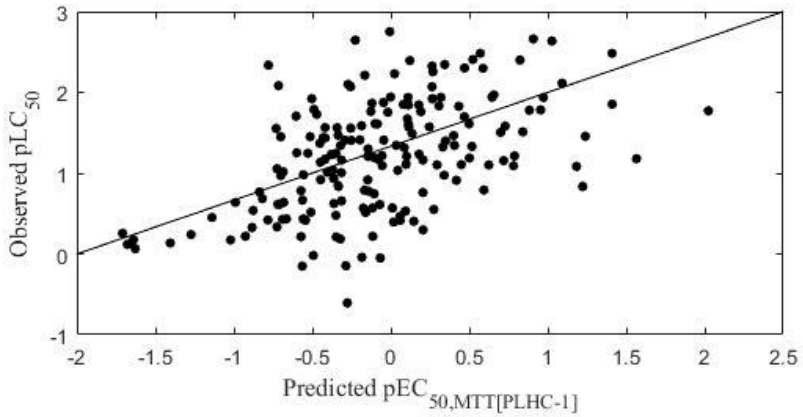


Figure 4.25. Plot of observed pLC_{50} versus predicted $pEC_{50,MTT[PLHC-1]}$ values.

4.6. Comparison of All QSTR Models

Fit, internal and external validation parameters for all of the generated QSTR models are compared in Table 4.40. When all QSTR models were compared in terms of fit and internal validation parameters, $pEC_{50,NR[PLHC-1]}$ model had the highest R^2, R^2_{adj} and CCC_{Tr} values whereas $pEC_{50,NR[GFS]}$ model showed highest F and Q^2_{LOO} values. Considering external validation parameters $pEC_{50,NR[GFS]}$ model were superior to the other QSTR models with highest R^2_{TEST}, Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{TEST} values. Furthermore, $pEC_{50,MTT[PLHC-1]}$ model showed the lowest $RMSE_{TR}$ value while $pEC_{50,NR[FHM]}$ model showed the lowest $RMSE_{TEST}$ and MAE_{TEST} values.

In order to compare predicted endpoint values from each of five models, correlation analyses were performed. The Pearson correlation coefficients (r), two-tailed significance values (p) and the number of chemicals used in the relevant analyses were summarized in Table 4.41.

It can be seen from Table 4.41 that strong correlations between $pEC_{50,EROD[PLHC-1]}$ - $pEC_{50,NR[GFS]}$; $pEC_{50,NR[PLHC-1]}$ - $pEC_{50,MTT[PLHC-1]}$ and $pEC_{50,NR[GFS]}$ - $pEC_{50,NR[FHM]}$ endpoints were found. However, the weak/moderate correlations shown in Table 4.41 can be attributed to the differences in the endpoint (i.e. $pEC_{50,EROD}$, $pEC_{50,NR}$ or $pEC_{50,MTT}$) and/or target organ or cell line that the endpoint was measured (i.e. PLHC-1, GFS or FHM).

Strong correlations between log K_{ow} and each of $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[GFS]}$ values were found, whereas a moderate and a weak correlations were found between log K_{ow} and $pEC_{50,NR[PLHC-1]}$, and $pEC_{50,MTT[PLHC-1]}$ endpoints, respectively. Moreover, the $pEC_{50,NR[FHM]}$ and log K_{ow} relationship showed no statistical significance.

For the predicted *in vitro* and experimental *in vivo* correlation analyses, strong correlations of $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[GFS]}$ values with *in vivo* fish lethality (pLC_{50}) data was found, whilst moderate correlations of $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ values with pLC_{50} data were observed. However, no statistically significant correlation was obtained between the $pEC_{50,NR[FHM]}$ values and pLC_{50} data.

The moderate correlations of each of $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ data with pLC_{50} data might be due to the differences in fish species used for the *in vivo* toxicity and *in vitro* cytotoxicity

determinations. *In vivo* pLC_{50} values were measured for *Oryzias latipes* while *in vivo* data were measured for *Poeciliopsis lucida* (hepatoma cell line) with two assays, NR and MTT.

When the AD of QSTR models generated in the present study are compared, the $pEC_{50,NR[PLHC-1]}$ model had the most broad AD. There were only 6 external set chemicals having hat values higher than the model's critical hat value (h^*=0.308) and predicted $pEC_{50,NR[PLHC-1]}$ values of 3 chemicals are outside of the predicted cytotoxicity range of the model. QSTR model for $pEC_{50,MTT[PLHC-1]}$ had also a similar performance with 11 external set chemicals outside the ADs of the model.

Additionally, for $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ models, the most important descriptor in the model equations was common while the second descriptors were from the same descriptor block. Indeed, the two data sets included common chemicals for which $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ values were obtained from the same source. Therefore, a correlation analysis was carried out between the experimental $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ values. A strong correlation was observed between $pEC_{50,\ NR[PLHC-1]}$ and $pEC_{50,\ MTT[PLHC-1]}$ values (n = 41; r = 0.899; p <0.01). Considering this and the strong correlation between predicted $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ values (n = 637; r = 0.833; p <0.01), it can be claimed that cytotoxicity values measured with NR and MTT assays strongly correlate with each other. This statement is in line with the previous studies reporting a strong correlation of NR and MTT dependent cytotoxicity values (Brüschweiler et al., 1995; Caminada et al., 2006).

On the other hand, $pEC_{50,EROD[PLHC-1]}$ model had 92.1% structural coverage for an external set of 353 chemicals. For this endpoint, a revised version of the external set used for $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ models was used. Considering the homogeneity in the $pEC_{50,EROD[PLHC-1]}$ dataset which was composed only of PAHs, NPAHs and azaarenes, it was expected that the model would have a better external performance for chemicals having structural similarity with the dataset chemicals. In line with what we expected, most of the APIs, pesticides and PAHs in the external set were within the AD of $pEC_{50,EROD[PLHC-1]}$ model. Therefore, the model was successful at predicting $pEC_{50,EROD[PLHC-1]}$ values for structurally similar chemicals in the external set.

Finally, $pEC_{50,NR[FHM]}$ model had a 89.1% structural coverage for 230 external set chemicals. For this endpoint, a limited external set was used compared to those of other four models. The cytotoxicity range of the experimental $pEC_{50,NR[FHM]}$ data is narrow. However, $pEC_{50,NR[FHM]}$ model was a validated QSTR model and had a high structural coverage of external set chemicals, hence it can be

effectively used for the cytotoxicity prediction of chemicals which fall in the AD of the model. Moreover, unlike the other $pEC_{50,NR}$ models, no statistically significant correlation was found between predicted $pEC_{50,NR[FHM]}$ values and neither log K_{ow} nor *in vivo* pLC_{50} values.

The most cytotoxic chemicals screened with the generated QSTR models were compared in Table 4.42. Regarding the $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions, most toxic chemicals were particularly from pharmaceuticals and pesticides. Therefore, the importance of assessment of the toxic potencies for these type of chemicals can be suggested.

It can be seen from Table 4.41 that tetrabromobisphenol A was found as one of the most cytotoxic chemicals depending on predictions of $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ models. Moreover, pyrazosulfuron ethyl was included in the most cytotoxic ten chemicals regarding the predictions of $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ models.

When chemicals with the highest $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions were compared, the same order was observed for the four common chemicals; tetrabromobisphenol A > glibenclamid > pyrazosulfuron ethyl > cyclosulfamuron. The order of $pEC_{50,NR[GFS]}$ cytotoxicity is tetrabromobisphenol A > cyclosulfamuron. Additionally, when the most toxic external set chemicals based on the prediction of $pEC_{50,NR[GFS]}$ model were screened, the relative toxicity order encountered for the three chemicals are the same; tetrabromobisphenol A > glibenclamid > pyrazosulfuron ethyl.

Three chemicals; tetrabromobisphenol A, 2,2-bis[4-(2-hydroxyethoxy)phenyl]propane and cyclosulfamuron were common among the most cytotoxic chemicals from $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions and their relative cytotoxicity values were in the same order; tetrabromobisphenol A > 2,2-bis[4-(2-hydroxyethoxy)phenyl]propane > cyclosulfamuron. Moreover, the order of toxicity was the same (tetrabromobisphenol A > 2,2-bis[4-(2-hydroxyethoxy)phenyl]propane) regarding *in vivo* pLC_{50} values of these two chemicals.

Table 4.40. Descriptors, fitting, internal and external validation parameters for the generated five QSTR models.

Fit and Internal Validation Parameters											
Model	Descriptors	R^2	R^2_{adj}	$RMSE_{Tr}$	CCC_{Tr}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}	R^2_{Yscr}	Q^2_{Yscr}
$pEC_{50,EROD[PLHC-1]}$	ATS3m	0.727	0.713	0.834	0.842	53.121	0.664	0.924	0.806	0.050	-0.157
$pEC_{50,NR[PLHC-1]}$	NaasC CATS2D_01_DN DLS_04	0.825	0.810	0.466	0.904	55.148	0.782	0.520	0.882	0.078	-0.149
$pEC_{50,NR[GFS]}$	GATS1v DP07	0.813	0.809	0.620	0.897	169.854	0.799	0.644	0.889	0.026	-0.051
$pEC_{50,NR[FHM]}$	X5v R3m+ GATS2v	0.627	0.580	0.538	0.770	13.417	0.502	0.621	0.689	0.114	-0.242
$pEC_{50,MTT[PLHC-1]}$	NaasC CATS2D_01_AN Mor28e	0.775	0.751	0.460	0.874	32.222	0.702	0.529	0.831	0.098	-0.192

External Validation Parameters											
Model	R^2_{TEST}	MAE_{TEST}	$RMSE_{TEST}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	$r^2_{m\ aver.}$	Δr^2_m	k'	k
$pEC_{50,EROD[PLHC-1]}$	0.856	0.329	0.380	0.851	0.850	0.943	0.914	0.736	0.139	0.973	1.014
$pEC_{50,NR[PLHC-1]}$	0.799	0.323	0.360	0.800	0.782	0.896	0.892	0.757	0.074	0.898	0.926
$pEC_{50,NR[GFS]}$	0.977	0.188	0.226	0.975	0.975	0.975	0.988	0.968	0.002	1.013	0.964
$pEC_{50,NR[FHM]}$	0.893	0.127	0.155	0.889	0.884	0.969	0.938	0.844	0.097	0.967	1.026
$pEC_{50,MTT[PLHC-1]}$	0.864	0.265	0.290	0.878	0.864	0.911	0.926	0.829	0.055	0.887	1.003

Table 4.41. Correlation between the predicted cytotoxicity values from each of the $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$, $pEC_{50,NR[FHM]}$, $pEC_{50,MTT[PLHC-1]}$ models.

Endpoint		$pEC_{50,NR[PLHC-1]}$	$pEC_{50,NR[GFS]}$	$pEC_{50,NR[FHM]}$	$pEC_{50,MTT[PLHC-1]}$
$pEC_{50,EROD[PLHC-1]}$	*r*	0.242*	**0.579***	0.923	0.236*
	p	0.000	0.000	0.251	0.000
	n	328	308	3	323
$pEC_{50,NR[PLHC-1]}$	*r*		**0.486***	0.205*	**0.833***
	p		0.000	0.005	0.000
	n		618	187	637
$pEC_{50,NR[GFS]}$	*r*			**0.615***	**0.466***
	p			0.000	0.000
	n			191	610
$pEC_{50,NR[FHM]}$	*r*				0.276*
	p				0.000
	n				187

*Correlation is significant at the 0.01 level (two-tailed).

Table 4.42. The most cytotoxic chemicals[a] predicted by the generated five QSTR models.

Chemical[b]	$pEC_{50,EROD[PLHC-1]}$ (mM)	Chemical[b]	$pEC_{50,NR[PLHC-1]}$ (mM)	Chemical[b]	$pEC_{50,NR[GFS]}$ (mM)
Tetrachloroguaiacol	5.833	Trichlorosyringol	2.39	Methotrexate	2.335
Tetrachlorocatechol	5.686	***Tetrabromobisphenol A***	2.343	***Tetrabromobisphenol A***	2.31
Testosterone propionate	5.666	Novobiocin	2.159	Cephalexin	2.308
Finasteride	5.642	Glibenclamide	1.997	Oxytetracycline	2.307
Pyrazosulfuron ethyl	5.549	Metipranolol	1.997	Cyclosulfamuron	2.291
2,4,6-tribromophenol	5.539	Pendimethalin	1.997	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	2.289
Camptothecin	5.495	Midazolam	1.951	Cefurexime axetil	2.28
Tetrachlorohydroquinone	5.495	Ketoconazole	1.789	Indomethacin	2.279
Nicosulfuron	5.485	**Pyrazosulfuron ethyl**	1.789	Bensulfuron-methyl	2.25
Methotrexate	5.426	*Cyclosulfamuron*	1.789	Acebutolol	2.246

Chemical[b]	$pEC_{50,NR[FHM]}$ (mM)	Chemical[b]	$pEC_{50,MTT[PLHC-1]}$ (mM)
2-chlorobenzyl chloride	-0.089	***Tetrabromobisphenol A***	2.023
Octanoic acid	-0.179	*Glibenclamide*	1.823
2-chloronitrobenzene	-0.186	**Pyrazosulfuron ethyl**	1.758
1,3-dibromopropane	-0.264	Prazosin	1.756
Glyoxylic acid	-0.293	Trimethoprim	1.621
o-chlorobenzonitrile	-0.374	*Cyclosulfamuron*	1.594
2,6-dichlorobenzonitrile	-0.394	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	1.565
Octanedinitrile	-0.427	Quinidine	1.552
3-chloronitrobenzene	-0.463	Terazosin	1.437
Dibutylamine	-0.478	2-(1,1-dimethyl)-4,6-dimethylphenol	1.411

[a]Based on the predicted values from Eq. 4.1 for $pEC_{50,EROD[PLHC-1]}$, Eq. 4.3 for $pEC_{50,NR[PLHC-1]}$, Eq. 4.7 for $pEC_{50,NR[GFS]}$, Eq. 4.10 for $pEC_{50,NR[FHM]}$ and Eq. 4.13 for $pEC_{50,MTT[PLHC-1]}$. [b]Common chemicals for $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions in bold and italics. Common chemicals for $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions in bold and underlined. Common chemicals for $pEC_{50,NR[PLHC-1]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions in italics. Common chemicals for $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions underlined.

6. CONCLUSIONS

In the present study, five QSTR/*in silico* models were generated for three different *in vitro* endpoints; $pEC_{50,EROD}$, $pEC_{50,NR}$ and $pEC_{50,MTT}$. All models were validated internally and externally and were found to be compatible with the Organization of Economic Co-operation Development (OECD) principles.

QSTR model for $pEC_{50,EROD[PLHC-1]}$ had one descriptor which is based on 2D-topological distance information of the molecules. This model was generated to explain the EROD induction potency of structurally similar compounds; mostly PAHs, NPAHs and azaarenes. Only TCDD and two PCBs provided a kind of variance in the data set. That fact might be a probable explanation of how such a complex enzymatic process can be explained with only one descriptor. In a QSTR study, molecular structure is the key point to build models explaining a target endpoint. Hence, if high structural similarity compounds present in a dataset to be modeled, then it would be relatively easier to explain their activity with one descriptor. The structural homogeneity of the dataset chemiclas was also considered while the predictive ability of QSTR model for this endpoint was being tested. The generated $pEC_{50,EROD[PLHC-1]}$ model was externally tested with 353 chemicals composed of the APIs, pesticides and PAHs. The $pEC_{50,EROD[PLHC-1]}$ model was successful at predicting $pEC_{50,EROD[PLHC-1]}$ values of 325 over 353 chemicals. Hence the structural coverage of $pEC_{50,EROD[PLHC-1]}$ model is 92.1%.

On the other hand, three QSTR models for $pEC_{50,NR}$ endpoint were generated; $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,NR[FHM]}$ models. QSTR model for $pEC_{50,NR[PLHC-1]}$ had 3 descriptors to explain cytotoxicity measured through NR assay. Molecular descriptors appeared in the model encoded information about electronic distribution and topological accessibility of selected atoms and specific position and charge of hydrogen-bond donors in a molecule. Also, drug-likeness was found to be effective in explaining $pEC_{50,NR[PLHC-1]}$ endpoint along with the relevant physicochemical properties used to encode this information. The proposed model was able to predict $pEC_{50,NR[PLHC-1]}$ values for 651 chemicals from various classes and had 98.6% structural coverage.

QSTR model for $pEC_{50,NR[GFS]}$ had 2 descriptors that are related to molecular profiles obtained through geometric distance matrices and with the 2D structure of molecules. When compared with the $pEC_{50,NR[PLHC-1]}$ model, $pEC_{50,NR[GFS]}$ model was able to explain the same NR uptake endpoint with fewer descriptors for a larger dataset. Even though $pEC_{50,NR[GFS]}$ dataset was larger and

composed of compounds from various classes, the $pEC_{50,NR[PLHC-1]}$ dataset include compounds with more complex molecular structures such as pharmaceuticals with 2 or 3-fused rings, complex pesticides etc. In order to represent such complexity in the dataset, the generated model included descriptors that are complex and dependent upon more parameters.

The QSTR model for $pEC_{50,NR[FHM]}$ had 3 descriptors encoding information about heteroatom content and 2D/3D-structure of molecules. One of the descriptors, GATS2v, appeared in $pEC_{50,NR[FHM]}$ model, was also present in $pEC_{50,NR[GFS]}$ model with only a difference in the target atom used to calculate the descriptor. This result supported the usage of those descriptors to explain neutral red uptake endpoint reported in the present study. The proposed model had an acceptable statistical performance yet it had two structural outliers in the AD. To test external predictivity of the generated $pEC_{50,NR[FHM]}$ model an external set of 230 chemicals was used. Of the 230 chemicals, the model could predict $pEC_{50,NR[FHM]}$ values for 205 chemicals resulting in 89.1% structural coverage.

Finally, QSTR model for $pEC_{50,MTT[PLHC-1]}$ held 3 descriptors encoding information related to electronic and topological information about selected atoms; specific position and charge of hydrogen-bond acceptors in a molecule and electronegativity of the molecule to explain the $pEC_{50,MTT[PLHC-1]}$ endpoint. The model could predict 649 chemicals with a structural coverage of 98.3%.

It was stated before that QSTR modelling on $pEC_{50,EROD}$, $pEC_{50,NR}$ and $pEC_{50,MTT}$ endpoints are limited in the literature. Therefore, this study could be an important source as it provides QSTR models for all the endpoints mentioned. Additionally, the generated models were able to predict cytotoxicity for a broad range of chemicals with no experimental data. $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ models could predict more than 600 diverse external set chemicals. Moreover, $pEC_{50,EROD[PLHC-1]}$ and $pEC_{50,NR[GFS]}$ models predicted more than 300 and 200 external set chemicals, respectively.

Based on the predicted cytotoxicity values, the most and least cytotoxic chemicals for each endpoint in the present study were screened. Among the studied diverse external set chemicals, particularly pharmaceuticals and pesticides were the most cytotoxic compounds depending on their $pEC_{50,EROD[PLHC-1]}$, $pEC_{50,NR[PLHC-1]}$, $pEC_{50,NR[GFS]}$ and $pEC_{50,MTT[PLHC-1]}$ predictions. The reported most and least cytotoxic chemicals were useful to provide an initial prioritization and to identify the need for further testing.

Considering moderate/strong correlations between *in vivo* toxicity and *in vitro* cytotoxicity reported in the present study, it can be suggested that the predicted cytotoxicity values by the models generated in this study are not only useful in terms of filling data gap but they can also be used for regulatory purposes where *in vivo* data are needed.

REFERENCES

Adhikari, N., Halder, A. K., Saha, A., Saha, K. D., Jha, T., 2015. Structural findings of phenylindoles as cytotoxic antimitotic agents in human breast cancer cell lines through multiple validated QSAR studies. Toxicology in vitro, 29, 1392-1404.

Aherne, G. W., English, J., Marks, V. 1985. The role of immunoassay in the analysis of microcontaminanls in river samples. Ecotoxicology and Environmental Safety, 9, 79-83.

Amin, A., Adhikari, N., Gayen, S., Jha, T., 2016. Insight into the Structural Requirements of Theophylline-Based Aldehyde Dehydrogenase lAl (ALDHlAl) Inhibitors Through Multi-QSAR Modeling and Molecular Docking Approaches. Current Drug Discovery Technologies, 13, 84-100.

Asensio, V., Kille, P., Morgan, A. J., Soto, M., Marigomez, I., 2007. Metallothionein expression and Neutral Red uptake as biomarkers of metal exposure and effect in Eisenia fetida and Lumbricus terrestris exposed to Cd. European Journal of Soil Biology, 43, 233-S238.

Armitage, J. M., Wania, F., Arnot, J. A., 2014. Application of mass balance models and the chemical activity concept to facilitate the use of *in vitro* toxicity data for risk assessment. Environmental Science and Technology, 48, 9770-9779.

Babich, H., Puerner, J. A., Borenfreund, E. 1986. *In vitro* cytotoxicity of metals to bluegill (BF-2) cells. Archives of Environmental Contamination and Toxicology, 15, 31-37.

Babich, H., Borenfreund, E. 1987. Structure-activity relationship (SAR) models established *in vitro* with the neutral red cytotoxicity assay. Toxicology in vitro, 1, 3-9.

Babich, H., Rosenberg, D. W., Borenfreund, E., 1991. *In vitro* cytotoxicity studies with the fish hepatoma cell line, PLHC-1 (*Poeciliopsis lucida*). Ecotoxicology and environmental safety, 21, 327-336.

Balch, G. C., Metcalfe, C. D., Huestis S. Y., 1995. Identification of potential fish carcinogens in sediment from Hamilton Harbour, Ontario, Canada. Environmental Toxicology and Chemistry, 14, 79–91.

Bandowe, B. A. M., Meusel, H., 2017. Nitrated polycyclic aromatic hydrocarbons (nitro-PAHs) in the environment–A review. Science of the Total Environment, 581, 237-257.

Bartram, A. E., Winter, M. J., Huggett, D. B., McCormack, P., Constantine, L. A., Hetheridge, M. J., Hutchinson, T. H., Kinter L. B., Ericson J. F., Sumpter J. P., Owen, S. F., 2012. *In vivo* and *in vitro* liver and gill EROD activity in rainbow trout (*Oncorhynchus mykiss*) exposed to the beta-blocker propranolol. Environmental Toxicology, 27, 573-582.

Bols, N. C., Dayeh, V. R., Lee, L. E. J., Schirmer, K. 2005. Use of fish cell lines in the toxicology and ecotoxicology of fish. Piscine cell lines in environmental toxicology. Biochemistry and Molecular Biology of Fishes, 6, 43-84.

Borenfreund, E., Shopsis, C., 1985. Toxicity monitored with a correlated set of cell-culture assays. Xenobiotica, 15, 705-711.

Borenfreund, E., Babich, H., Martin-Alguacil, N., 1988. Comparisons of two *in vitro* cytotoxicity assays-the neutral red (NR) and tetrazolium MTT tests. Toxicology in vitro, 2, 1-6.

Brandão, J. C., Bohets, H. H., Van De Vyver, I. E., Dierickx, P. J., 1992. Correlation between the *in vitro* cytotoxicity to cultured fathead minnow fish cells and fish lethality data for 50 chemicals. Chemosphere, 25, 553-562.

Broto, P., Moreau, G., Vandycke, C. 1984. Molecular structures: perception, autocorrelation descriptor and sar studies: system of atomic contributions for the calculation of the n-octanol/water partition coefficients. European Journal of Medicinal Chemistry, 19, 71-78.

Bruce, R., Santodonato, J., Neal, M., 1987. Summary review of the health effects associated with phenol. Toxicology and Industrial Health, 3, 535.

Brüschweiler, B. J., Würgler, F. E., Fent, K., 1995. Cytotoxicity *in vitro* of organotin compounds to fish hepatoma cells PLHC-1 (*Poeciliopsis lucida*). Aquatic Toxicology, 32, 143-160.

Bury, N. R., Schnell, S., Hogstrand, C., 2014. Gill cell culture systems as models for aquatic environmental monitoring. The Journal of Experimental Biology, 217, 639-650.

Caballero, J., Fernández, M., 2006. Linear and nonlinear modeling of antifungal activity of some heterocyclic ring derivatives using multiple linear regression and Bayesian-regularized neural networks. Journal of Molecular Modeling, 12, 168-181.

Caminada, D., Escher, C., Fent, K., 2006. Cytotoxicity of pharmaceuticals found in aquatic systems: comparison of PLHC-1 and RTG-2 fish cell lines. Aquatic Toxicology, 79, 114-123.

Chen, G., Zheng, S., Luo, X., Shen, J., Zhu, W., Liu, H., Gui, C., Zhang, J., Zheng, M., Puah, C. M., Chen, K., Jiang, H., 2005. Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score. Journal of Combinatorial Chemistry, 7, 398-406.

Chirico, N., Gramatica, P., 2011. Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. Journal of Chemical Information and Modeling, 51, 2320-2335.

Chirico, N., Gramatica, P., 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. Journal of Chemical Information and Modeling, 52, 2044-2058.

Collins, J., Brown, J., Alexeeff, G., Salmon, A., 1998. Potency equivalency factors for some polycyclic aromatic hydrocarbons and polycyclic aromatic hydrocarbon derivatives. Regulatory Toxicology and Pharmacology. 28, 45–54.

Consonni, V., Ballabio, D., Todeschini, R., 2009. Comments on the definition of the Q^2 parameter for QSAR validation. Journal of chemical information and modeling, 49, 1669-1678.

Consonni, V., Ballabio, D., Todeschini, R. 2010. Evaluation of model predictive ability by external validation techniques. Journal of chemometrics, 24, 194-201.

Contrera, J. F., Matthews, E. J., Benz, R. D., 2003. Predicting the carcinogenic potential of pharmaceuticals in rodents using molecular structural similarity and E-state indices. Regulatory Toxicology and Pharmacology, 38, 243-259.

Chirico, N., Gramatica, P., 2012. Real External Predictivity of QSAR Models. Part 2. New inter-comparable thresholds for different validation criteria and the need for scatter plot inspection. Journal of Chemical Information and Modelling, 52, 2044–2058.

DRAGON v. 6, 2013. Talete srl, Milan, Italy. http://www.talete.mi.it/ Date accessed July 2018.

Duchowicz, P. R., Vitale, M. G., Castro, E. A., Fernández, M., Caballero, J., 2007. QSAR analysis for heterocyclic antifungals. Bioorganic & medicinal chemistry, 15, 2680-2689.

ECHA, 2006. 2',6,6'-tetrabromo-4,4'-isopropylidenediphenol (tetrabromobisphenol-A or TBBP-A). https://echa.europa.eu/documents/10162/32b000fe-b4fe-4828-b3d3-93c24c1cdd51 Date accessed July 2018.

Fent, K., Hunn, J., 1996. Cytotoxicity of organic environmental chemicals to fish liver cells (PLHC-1). Marine Environmental Research, 42, 377-382.

Fent, K., Bätscher, R., 2000. Cytochrome P4501A induction potencies of polycyclic aromatic hydrocarbons in a fish hepatoma cell line: demonstration of additive interactions. Environmental Toxicology and Chemistry, 19, 2047-2058.

Fereidoonnezhad, M., Mostoufi, A., Aliyan, F., 2017. N-Phenyl Ureidobenzenesulfonate derivatives as novel anticancer agents: QSAR and Molecular docking studies. Trends in Pharmaceutical Sciences, 3, 83-104.

Fotakis, G., Timbrell, J. A., 2006. *In vitro* cytotoxicity assays: comparison of LDH, neutral red, MTT and protein assay in hepatoma cell lines following exposure to cadmium chloride. Toxicology Letters, 160, 171-177.

Gaines, T. B., 1969. Acute toxicity of pesticides. Toxicology and Applied Pharmacology, 14, 515-534.

Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. Journal of Computational Chemistry, 34, 2121-2132.

Gramatica, P., Cassani, S., Chirico, N., 2014. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. Journal of Computational Chemistry, 35, 1036-1044.

Gravell, M., Malsberger, R. G., 1965. A permanent cell line from the fathead minnow (*Pimephales promelas*). Annals of the New York Academy of Sciences, 126, 555-565.

Golbraikh, A., Tropsha, A., 2002. Beware of q^2!. Journal of Molecular Graphics and Modelling, 20, 269-276.

Gräns, J., Johansson, J., Michelová, M., Wassmur, B., Norström, E., Wallin, M., Celander, M. C., 2015. Mixture effects between different azoles and β-naphthoflavone on the CYP1A biomarker in a fish cell line. Aquatic Toxicology, 164, 43-51.

Hahn, M. E., Woodward, B. L., Stegeman, J. J., Kennedy, S. W., 1996. Rapid assessment of induced cytochrome P4501 a protein and catalytic activity in fish hepatoma cells grown in multiwell plates: Response to TCDD, TCDF, and two planar PCBS. Environmental Toxicology and Chemistry, 15, 582-591.

Hall, L. H., Kier, L. B., 1995. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. Journal of Chemical Information and Computer Sciences, 35, 1039-1045.

Hall, L. H., 2004. A Structure-Information Approach to the Prediction of Biological Activities and Properties. Chemistry & Biodiversity, 1, 183-201.

Henschel, K. P., Wenzel, A., Diedrich, M., Fliedner, A., 1997. Environmental hazard assessment of pharmaceuticals. Regulatory Toxicology and Pharmacology, 25, 220-225.

IARC, 1972. International Agency for Research on Cancer, Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man. World Health Organization, Geneva.

IBM SPSS v. 24, IBM Corp, Armonk, NY. https://www.ibm.com/us-en/marketplace?lnk=mp Date accessed July 2018.

Jung, D. K., Klaus, T., Fent, K., 2001. Cytochrome P450 induction by nitrated polycyclic aromatic hydrocarbons, azaarenes, and binary mixtures in fish hepatoma cell line PLHC-1. Environmental Toxicology and Chemistry, 20, 149-159.

Keller, H. R., Massart, D. L., Brans, J. P., 1991. Multicriteria decision making: a case study. Chemometrics and Intelligent Laboratory Systems, 11, 175-189.

Kier, L. B., Hall, L. H., 1981. Derivation and significance of valence molecular connectivity. Journal of Pharmaceutical Sciences, 70, 583-589.

Knauer, K., Lampert, C., Gonzalez–Valero, J., 2007. Comparison of *in vitro* and *in vivo* acute fish toxicity in relation to toxicant mode of action. Chemosphere, 68, 1435–1441.

Laville, N., Ait-Aıssa, S., Gomez, E., Casellas, C., Porcher, J. M., 2004. Effects of human pharmaceuticals on cytotoxicity, EROD activity and ROS production in fish hepatocytes. Toxicology, 196, 41-55.

Lin, L. I., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45, 255–268.

Lin, L. I., 1992. Assay validation using the concordance correlation coefficient. Biometrics, 48, 599–604.

LoPachin, R. M., Gavin, T., Geohagen, B. C., Das, S., 2007. Neurotoxic mechanisms of electrophilic type-2 alkenes: soft–soft interactions described by quantum mechanical parameters. Toxicological Sciences, 98, 561-570.

Marks, D. C., Belov, L., Davey, M.W., Davey, R.A., Kidman, A.D., 1992. The MTT cell viability assay for cytotoxicity testing in multidrug-resistant human leukemic cells. Leukemia Research, 16, 1165-1173.

Michalowicz, J., Duda, W., 2004. Chlorophenols and their derivatives in waters of the drainage of the Dzierzazna river State and anthropogenic changes of the quality of waters in Poland. Hydrological Committee of Polish Geographical Society, University of Lodz, Lodz.

Michałowicz, J., 2005. The occurrence of chlorophenols, chlorocatechols and chlorinated methoxyphenols in drinking water of the largest cities in Poland. Polish Journal of Environmental Studies, 14, 327–333.

Michałowicz, J., Duda, W., 2007. Phenols-Sources and Toxicity. Polish Journal of Environmental Studies, 16, 347–362.

Moreau, G., Broto, P., 1980a. Autocorrelation of molecular structures. Application to SAR studies. Nouveau Journal De Chimie-New Journal of Chemistry, 4, 757–764.

Moreau, G., Broto, P., 1980b. The autocorrelation of a topological structure: a new molecular descriptor. Nouveau Journal De Chimie-New Journal of Chemistry, 4, 359–360.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. Journal of Immunological Methods, 65, 55-63.

Nagelkerke, N. J., 1991. A note on a general definition of the coefficient of determination. Biometrika, 78, 691-692.

Nathalie, D., Yannick, G., Caroline, B., Sandrine, D., Claude, F., Corinne, C., Pierre-Jacques, F., 2006. Assessment of the phototoxic hazard of some essential oils using modified 3T3 neutral red uptake assay. Toxicology in Vitro, 20, 480-489.

OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models. Environment Health and Safety Publications. Series on Testing and Assessment No. 69. Paris, France.

OECD, 2003. SIDS Initial Assessment Report For SIAM 17: Benzene, 1-chloro-2-(chloromethyl). Paris, France

Ojha, P. K., Mitra, I., Das, R. N., Roy, K. 2011. Further exploring r_m^2 metrics for validation of QSPR models. Chemometrics and Intelligent Laboratory Systems, 107, 194-205.

Önlü, S., Saçan, M. T. 2017. An in silico algal toxicity model with a wide applicability potential for industrial chemicals and pharmaceuticals. Environmental Toxicology and Chemistry, 36, 1012-1019.

Paasivirta, J., 1995. Chemical Ecotoxicology. CRC Press, Boca Raton, Florida.

Papa, E., Battaini, F., Gramatica, P., 2005. Ranking of aquatic toxicity of esters modelled by QSAR. Chemosphere, 58, 559-570.

Pearlman, R. S., Yalkowsky, S. H., Banerjee, S., 1984. Water solubilities of polynuclear aromatic and heteroaromatic compounds. Joutnal of Physical and Chemical Reference Data, 13, 555–562.

Repetto, G., Del Peso, A., Zurita, J. L., 2008. Neutral red uptake assay for the estimation of cell viability/cytotoxicity. Nature Protocols, 3, 1125.

Results of aquatic toxicity tests of chemicals conducted by Ministry of the Environment in Japan. http://www.env.go.jp/en/chemi/sesaku/aquatic_Mar_2018.pdf Date accessed in July 2018.

Richard, A. M., 2006. Future of toxicology--predictive toxicology: An expanded view of "chemical toxicity". Chemical Research in Toxicology, 19, 1257-1262.

Rogers, I. H., Levings, C. D., Lockhart, W. L., Norstrom, R. J., 1989. Observations on overwintering juvenile chinook salmon (oncorhynchus tsawytscha) exposed to bleached kraft mill effluent in the upper fraser river, British Columbia. Chemosphere, 19, 1853–1868, 1989.

Roig, B., Marquenet, B., Delpla, I., Bessonneau, V., Sellier, A., Leder, C., Thomas O., Bolek R., Kummerer, K., 2014. Monitoring of methotrexate chlorination in water. Water Research, 57, 67-75.

Roy, K., Das, R. D., Ambure, P., Aher. R. B., 2016. Be Aware of Error Measures. Further Studies on Validation of Predictive QSAR Models. Chemometrics and Intelligent Laboratory Systems, 152, 18–33.

Saito, H., Koyasu, J., Yoshida, K., Shigeoka, T., Koike, S., 1993. Cytotoxicity of 109 chemicals to goldfish GFS cells and relationships with 1-octanol/water partition coefficients. Chemosphere, 26, 1015-1028.

Sanborn, M. D., Cole, D., Abelsohn, A., Weir, E., 2002. Identifying and managing adverse environmental health effects: 4. Pesticides. Canadian Medical Association Journal, 166, 11, 1431-1436.

Schuur, J. H., Selzer, P., Gasteiger, J., 1996. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. Journal of Chemical Information and Computer Sciences, 36, 334-344.

Schüürmann, G., Ebert, R. U., Chen, J., Wang, B., Kühne, R., 2008. External validation and prediction employing the predictive squared correlation coefficient. Test set activity mean vs training set activity mean. Journal of Chemical Information and Modeling, 48, 2140-2145.

Schweigert, N., Alexander, J., Zehnder, J., Eggen, R., 2001. Chemical properties of catechols and their molecular modes of toxic action in cells from microoorganisms to mammals. Environmental Microbiology, 3, 81.

SPARTAN v. 10, 2011. Wavefunction Inc., Irvine, CA, USA. http://wavefun.com Date accessed July 2018.

Stadnicka-Michalak, J., Tanneberger, K., Schirmer, K., Ashauer, R., 2014. Measured and modeled toxicokinetics in cultured fish cells and application to *in vitro-in vivo* toxicity extrapolation. PloS One, 9.

Sullivan, K. M., Manuppello, J. R., Willett, C. E., 2014. Building on a solid foundation: SAR and QSAR as a fundamental strategy to reduce animal testing. SAR and QSAR in Environmental Research. 25, 357-365.

Todeschini, R., Consonni, V. 2008. Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, Germany.

Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics. Vol. 41, 2nd edition, Wiley-VCH, Weinheim, Germany.

Topliss, J. G., Costello, R. J., 1972. Chance correlations in structure-activity studies using multiple regression analysis. Journal of Medicinal Chemistry, 15, 1066-1068.

Tropsha, A., Gramatica, P., Gombar, V., 2003. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR Models. Molecular Informatics. 22, 69–77.

Tugcu, G., Saçan, M. T., Vračko, M., Novič, M., Minovski, N., 2012. QSTR modeling of the acute toxicity of pharmaceuticals to fish. SAR and QSAR in Environmental Research, 23, 297-310.

Ukelis, U., Kramer, P. J., Olejniczak, K., Mueller, S. O., 2008. Replacement of *in vivo* acute oral toxicity studies by *in vitro* cytotoxicity methods: Opportunities, limits and regulatory status. Regulatory Toxicology and Pharmacology, 51, 108-118.

US EPA, 2016. The United States Environmental Protection Agency Priority Pollutant List. https://www.epa.gov/sites/production/files/2015-09/documents/priority-pollutant-list-epa.pdf Date accessed July 2018.

Van de Loosdrecht, A. A., Beelen, R. H. J., Ossenkoppele, G. J., Broekhoven, M. G., Langenhuijsen, M. M. A. C., 1994. A tetrazolium-based colorimetric MTT assay to quantitate human monocyte mediated cytotoxicity against leukemic cells from cell lines and patients with acute myeloid leukemia. Journal of Immunological Methods, 174, 311-320.

Woodland, C. L. A., Maly, E. J., 1997. The effects of tetrachloroguaiacol on the growth, survival, and development of the fathead minnow, *Pimephales promelas*. Archives of Environmental Contamination and Toxicology, 32, 394-398.

APPENDIX A: EXTERNAL SET CHEMICALS FOR $pEC_{50,EROD[PLHC-1]}$ MODEL

Table A1. External set chemicals for $pEC_{50,EROD[PLHC-1]}$ model.

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h^*=0.273)	ATS3m	pLC_{50} (mM)
Acebutolol	37517-30-9	3.315	0.046	3.752	
Acridine	260-94-6	0.685	0.198	3.216	
Acrivastine	87848-99-5	3.261	0.046	3.741	
Acyclovir	59277-89-3	1.647	0.103	3.412	
Alprenolol	13655-52-2	1.514	0.113	3.385	
Amitriptyline	50-48-6	3.139	0.046	3.716	
Amobarbital	57-43-2	2.393	0.060	3.564	
Ampicillin	69-53-4	4.793	0.110	4.053	
Antipyrine	60-80-0	0.817	0.182	3.243	
Atovaquone	95233-18-4	5.092	0.136	4.114	
Bumetanide	28395-03-1	4.434	0.085	3.980	
Bupropion	34841-39-9	1.858	0.088	3.455	
Caffeine	58-08-2	2.039	0.077	3.492	
Camptothecin	7689-03-4	5.495	0.179	4.196	
Capecitabine	154361-50-9	4.552	0.093	4.004	
Carvedilol	72956-09-3	4.812	0.112	4.057	
Cefuroxime	55268-75-2	5.333	0.161	4.163	
Cephalexin	15686-71-2	4.753	0.107	4.045	
Chloramphenicol	56-75-7	3.939	0.060	3.879	
Chlorpromazine	50-53-3	4.341	0.079	3.961	
Chlorpropamide	94-20-2	3.154	0.046	3.719	
Cimetidine	51481-61-9	1.558	0.109	3.394	
Ciprofloxacin	85721-33-1	4.474	0.087	3.988	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h^*=0.273)	ATS3m	pLC_{50} (mM)
Clonazepam	1622-61-3	4.366	0.081	3.966	
Clonidine	4205-90-7	2.226	0.067	3.530	
Cytarabine	147-94-4	2.913	0.047	3.670	
Desipramine	50-47-5	3.100	0.046	3.708	
Estradiol	50-28-2	4.012	0.063	3.894	
Ethinylestradiol	57-63-6	4.567	0.094	4.007	
Famciclovir	104227-87-4	2.888	0.047	3.665	
Famotidine	76824-35-6	2.928	0.047	3.673	
Finasteride	98319-26-7	5.642	0.196	4.226	
Fleroxacin	79660-72-3	5.249	0.152	4.146	
Florfenicol	73231-34-2	4.572	0.094	4.008	
Flunitrazepam	1622-62-4	3.958	0.061	3.883	
Flutamide	13311-84-7	2.820	0.049	3.651	
Furaltadone	139-91-3	3.173	0.046	3.723	
Gemcitabine	95058-81-4	3.880	0.058	3.867	
Hydrochlorothiazide	58-93-5	4.744	0.106	4.043	
Imipramine	50-49-7	3.315	0.046	3.752	
Indomethacin	53-86-1	4.518	0.090	3.997	
Isotretinoin	4759-48-2	2.628	0.053	3.612	
Ketoprofen	22071-15-4	2.329	0.063	3.551	
Labetalol	36894-69-6	3.374	0.047	3.764	
Lamotrigine	84057-84-1	3.850	0.057	3.861	
Lansoprazole	103577-45-3	4.596	0.096	4.013	
Levofloxacin	100986-85-4	5.396	0.168	4.176	
Lidocaine	137-58-6	1.608	0.106	3.404	
Lomefloxacin	98079-51-7	5.102	0.137	4.116	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h*=0.273)	ATS3m	pLC_{50} (mM)
L-tryptophan	73-22-3	0.965	0.165	3.273	
Metaxalone	1665-48-1	0.783	0.186	3.236	
Methotrexate	59-05-2	5.426	0.171	4.182	
Metipranolol	22664-55-7	3.252	0.046	3.739	
Metoprolol	37350-58-6	1.612	0.105	3.405	
Midazolam	59467-70-8	4.582	0.095	4.010	
Nadolol	42200-33-9	3.512	0.048	3.792	
Naproxen	22204-53-1	1.927	0.084	3.469	
Nisoldipine	63675-72-9	4.999	0.128	4.095	
Norfloxacin	70458-96-7	4.297	0.077	3.952	
Ofloxacin	82419-36-1	5.396	0.168	4.176	
Oxprenolol	6452-71-7	1.887	0.086	3.461	
Oxyphenbutazone	129-20-4	4.066	0.065	3.905	
Paroxetine	61869-08-7	3.639	0.051	3.818	
Pentobarbital	76-74-4	2.727	0.050	3.632	
Phenobarbital	50-06-6	2.879	0.048	3.663	
Phenoxymethylpenicillinic acid	87-08-1	4.542	0.092	4.002	
Phenylbutazone	50-33-9	3.772	0.054	3.845	
Phenytoin	57-41-0	2.903	0.047	3.668	
Pindolol	13523-86-9	1.691	0.099	3.421	
Prazosin	19216-56-9	4.974	0.126	4.090	
Procaine	59-46-1	1.298	0.132	3.341	
Progesterone	57-83-0	4.945	0.123	4.084	
Promazine	58-40-2	3.723	0.053	3.835	
Quinidine	56-54-2	4.326	0.078	3.958	
Quinine	130-95-0	4.326	0.078	3.958	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h*=0.273)	ATS3m	pLC_{50} (mM)
Ranitidine	66357-35-5	2.599	0.054	3.606	
Risperidone	106266-06-2	5.357	0.163	4.168	
Sotalol	3930-20-9	2.020	0.078	3.488	
Sulfaguanidine	57-67-0	1.863	0.088	3.456	
Sulfaphenazole	526-08-9	3.885	0.058	3.868	
Sulfasalazine	599-79-1	5.038	0.131	4.103	
Sumatriptan	103628-46-2	2.874	0.048	3.662	
Terazosin	63590-64-7	5.063	0.134	4.108	
Terbinafine	91161-71-6	2.403	0.060	3.566	
Testosterone	58-22-0	4.611	0.097	4.016	
Theophylline	58-55-9	1.490	0.115	3.380	
Timolol	26839-75-8	3.154	0.046	3.719	
Tolazamide	1156-19-0	4.027	0.064	3.897	
Tolbutamide	64-77-7	2.864	0.048	3.660	
Trenbolone acetate	10161-34-9	4.420	0.084	3.977	
Triamterene	396-01-0	2.790	0.049	3.645	
Trifluprom azine	146-54-3	4.812	0.112	4.057	
Warfarin	81-81-2	3.782	0.055	3.847	
2,6-dichlorobenzonitrile	1194-65-6	1.087	0.152	3.298	
Acetochlor	34256-82-1	2.859	0.048	3.659	
Ametryn	834-12-8	1.102	0.151	3.301	
Anilofos	64249-01-0	4.803	0.111	4.055	
Benalaxyl	71626-11-4	3.865	0.057	3.864	
Benomyl	17804-35-2	2.830	0.048	3.653	
Bensulfuron-methyl	83055-99-6	5.018	0.130	4.099	
Beta-cyfluthrin	68359-37-5	4.906	0.120	4.076	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h*=0.273)	ATS3m	pLC_{50} (mM)
Bispyribac	125401-75-4	5.166	0.144	4.129	
Bromoxynil	1689-84-5	4.753	0.107	4.045	
Chloridazon	1698-60-8	2.781	0.049	3.643	
Chlorimuron-ethyl	90982-32-4	5.401	0.168	4.177	
Chlorothalonil	1897-45-6	5.426	0.171	4.182	
Chlorotoluron	15545-48-9	1.529	0.112	3.388	
Chlorsulfuron	64902-72-3	5.136	0.141	4.123	
Cyanazine	21725-46-2	1.637	0.103	3.410	
Diclofop methyl	51338-27-3	4.194	0.071	3.931	
Diclofop P	40843-25-2	3.958	0.061	3.883	
Dimethachlon	24096-53-5	2.354	0.062	3.556	
Ethalfluraline	55283-68-6	4.012	0.063	3.894	
Ethametsulfuron	111353-84-5	5.141	0.141	4.124	
Ethoxyquin	91-53-2	1.691	0.099	3.421	1.51
Fenamiphos	22224-92-6	3.870	0.058	3.865	
Fenhexamid	126833-17-8	4.999	0.128	4.095	
Fenoxaprop	95617-09-7	3.698	0.052	3.830	
Fenpropidin	67306-00-7	2.358	0.062	3.557	
Fluazifop P	83066-88-0	3.502	0.048	3.790	
Flumetsulam	98967-40-9	4.552	0.093	4.004	
Fluroxypyr	69377-81-7	3.949	0.060	3.881	
Iprodione	36734-19-7	4.057	0.065	3.903	
Cybutryne	28159-98-0	1.676	0.100	3.418	
Isoproturon	34123-59-6	0.793	0.185	3.238	
Mefenacet	73250-68-7	3.207	0.046	3.730	
Metalaxyl	57837-19-1	3.188	0.046	3.726	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
Methabenzthiazuron	18691-97-9	2.089	0.074	3.502	
Metolachlor	51218-45-2	3.301	0.046	3.749	
Metribuzin	21087-64-9	2.457	0.058	3.577	
Metsulfuron-methyl	74223-64-6	5.023	0.130	4.100	
Molinate	2212-67-1	0.754	0.189	3.230	
Nicosulfuron	111991-09-4	5.485	0.177	4.194	
Oxadiargyl	39807-15-3	4.724	0.105	4.039	
Oxadiazon	19666-30-9	4.773	0.109	4.049	
Oxadixyl	77732-09-3	3.109	0.046	3.710	
Pendimethalin	40487-42-1	3.266	0.046	3.742	
Phenmedipham	13684-63-4	2.535	0.055	3.593	
Pretilachlor	51218-49-6	3.453	0.047	3.780	2.11
Prometryn	7287-19-6	1.416	0.121	3.365	
Quinclorac	84087-01-4	2.884	0.048	3.664	
Quizalofop P	94051-08-8	4.027	0.064	3.897	
Tebuthiuron	34014-18-1	2.437	0.059	3.573	
Terbumeton	33693-04-8	1.028	0.158	3.286	
Terbuthylazine	5915-41-3	1.146	0.146	3.310	
Terbutryn	886-50-0	1.495	0.115	3.381	
Thiobencarb	28249-77-6	2.481	0.057	3.582	2.30
Thiophanate methyl	23564-05-8	3.379	0.047	3.765	
Tribenuron	106040-48-6	5.225	0.149	4.141	
1,1,2,2-tetrachloroethane	79-34-5	2.604	0.053	3.607	
1,2,3-trichloropropane	96-18-4	0.925	0.170	3.265	
Acetic acid, bromo-, 2-butene-1,4-diyl ester	20679-58-7	3.978	0.062	3.887	
Pentachloroethane	76-01-7	4.479	0.088	3.989	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
Tetrachloroethene	127-18-4	2.466	0.058	3.579	1.07
1-(*n*-phenylamino)-naphthalene	90-30-2	1.392	0.123	3.360	2.50
3-hydroxy-2-naphthoic acid	92-70-6	0.940	0.168	3.268	
Isopropylnaphthalene	29253-36-9	0.238	0.257	3.125	**2.36**
n-phenyl-2-naphthylamine	135-88-6	1.220	0.139	3.325	
1,2,3-trichlorobenzene	87-61-6	3.026	0.046	3.693	1.75
1,2-dichlorobenzene	95-50-1	0.837	0.180	3.247	1.59
1,4-benzenediamine, *n*,*n*'-*bis*(1-methylpropyl)-	101-96-2	0.812	0.182	3.242	**2.78**
2,4,6-trichlorophenylhydrazine	5329-12-4	2.344	0.062	3.554	2.33
2,4,6-trinitrotoluene	118-96-7	1.995	0.080	3.483	
2,4-diamino-6-nitrotoluene	6629-29-4	0.194	0.264	3.116	
2,4-dichlorotoluene	95-73-8	0.204	0.262	3.118	
2,4-dinitrotoluene	121-14-2	0.474	0.225	3.173	
2,5-dichlorotoluene	19398-61-9	0.204	0.262	3.118	**1.60**
2,6-diamino-4-nitrotoluene	59229-75-3	0.194	0.264	3.116	
2,6-dichlorotoluene	118-69-4	0.204	0.262	3.118	**1.85**
2,6-dinitrotoluene	606-20-2	0.675	0.199	3.214	0.73
2-amino-4,6-dinitrotoluene	35572-78-2	1.176	0.143	3.316	
2-chlorohydroquinonedimethylether	2100-42-7	0.651	0.202	3.209	0.79
2-methyl-4-chlorophenoxyacetic acid	94-74-6	0.695	0.196	3.218	
3,4-dichlorotoluene	95-75-0	1.289	0.133	3.339	1.57
3,5-bis(trifluoromethyl)benzylamine	85068-29-7	2.039	0.077	3.492	**1.34**
4,6-dichlororesorcinol	137-19-9	1.318	0.130	3.345	
4-nitrotoluene-2-sulphonic acid	121-03-9	1.833	0.090	3.450	
4-toluenesulfonyl chloride	98-59-9	0.577	0.211	3.194	0.39
Atenolol	29122-68-7	1.632	0.104	3.409	

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
Benzotrifluoride	98-08-8	0.651	0.202	3.209	1.01
Dimethyl phthalate	131-11-3	0.832	0.180	3.246	
Diuron	330-54-1	2.874	0.048	3.662	
Fenobucarb	3766-81-2	0.886	0.174	3.257	1.33
Metaxylene hexafluoride	402-31-3	1.254	0.136	3.332	1.47
o-toluenesulfonamide	88-19-7	0.459	0.227	3.170	
Oxyfluorfen	42874-03-3	4.577	0.094	4.009	
Propoxur	114-26-1	0.695	0.196	3.218	
Thiamphenicol	15318-45-3	4.528	0.091	3.999	
2-(1,1-dimethyl)-4,6-dimethylphenol	1879-09-0	0.788	0.185	3.237	1.85
2,3,4,5-tetrachlorophenol	4901-51-3	5.082	0.136	4.112	
2,3,5,6-tetrachlorophenol	935-95-5	4.690	0.102	4.032	
2,3,5-trichlorophenol	933-78-8	2.805	0.049	3.648	
2,3,6-trichlorophenol	933-75-5	3.242	0.046	3.737	
2,4,6-tribromophenol	118-79-6	5.539	0.184	4.205	**2.34**
2,4-dibromophenol	615-58-7	3.541	0.049	3.798	**1.84**
2,4-di-*tert*-butylphenol	96-76-4	1.524	0.112	3.387	2.48
2,6-di-*sec*-butylphenol	5510-99-6	1.431	0.120	3.368	3.14
3,4,5-trichloroguaiacol	57057-83-7	4.400	0.083	3.973	
3,4,5-trichlorophenol	609-19-8	3.365	0.046	3.762	
3-trifluoromethyl-4-nitrophenol	88-30-2	1.269	0.135	3.335	
4-(2,4-dichlorophenoxy)-phenol	40843-73-0	2.677	0.051	3.622	
4,5,6-trichloroguaiacol	2668-24-8	4.400	0.083	3.973	
4,5-dichloroguaiacol	2460-49-3	2.363	0.062	3.558	
4-chloro-2-nitrophenol	89-64-5	0.327	0.245	3.143	1.20
6-*tert*-butyl-*o*-cresol	2219-82-1	0.282	0.251	3.134	1.58

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h*=0.273)	ATS3m	pLC_{50} (mM)
Butylated hydroxyanisole	25013-16-5	0.557	0.214	3.190	**1.49**
Tetrachloroguaiacol	2539-17-5	5.833	0.220	4.265	
Trichlorosyringol	2539-26-6	5.254	0.152	4.147	
1-nitropyrene	5522-43-0	3.252	0.046	3.739	
2-acetamidophenoxazin-3-one	1916-55-8	2.766	0.050	3.640	
2-amino-7-methoxyphenoxazin-3-one		2.761	0.050	3.639	
2-aminophenoxazin-3-one	1916-59-2	2.074	0.075	3.499	
9-vinylcarbazole	1484-13-5	1.392	0.123	3.360	
Dibenzo[b,f]cyclohepten-1-one	2222-33-5	2.142	0.072	3.513	
Dibenzothiophene	132-65-0	1.372	0.125	3.356	2.12
Fluorene	86-73-7	0.160	0.269	3.109	
Phenanthrene	85-01-8	0.744	0.191	3.228	**2.10**
Phenothiazine	92-84-2	2.182	0.070	3.521	2.41
2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)	6864-37-5	2.147	0.071	3.514	**1.04**
Ethyl trichloroacetate	515-84-4	1.338	0.128	3.349	0.64
Testosterone propionate	57-85-2	5.666	0.199	4.231	
Tetracaine	94-24-6	1.627	0.104	3.408	
2,3,4-trichloroaniline	634-67-3	3.684	0.052	3.827	
2,3-dichloroaniline	608-27-5	1.833	0.090	3.450	
2,4,5-trichloroaniline	636-30-6	2.717	0.051	3.630	
2,4,6-trichloroaniline	634-93-5	2.000	0.079	3.484	1.57
2,4-dichloroaniline	554-00-7	0.346	0.242	3.147	1.30
2,5-dichloroaniline	95-82-9	0.346	0.242	3.147	1.87
2,6-dichloroaniline	608-31-1	0.965	0.165	3.273	
3,4,5-trichloroaniline	634-91-3	3.340	0.046	3.757	
3,4-dichloroaniline	95-76-1	1.323	0.130	3.346	1.17

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (h*=0.273)	ATS3m	pLC_{50} (mM)
4-chloro-2-nitroaniline	89-63-4	0.248	0.256	3.127	**1.01**
Chlorfluoroaniline	21397-08-0	0.621	0.206	3.203	
Dithiodianiline	722-27-0	3.355	0.046	3.760	
Dansylglycine	1091-85-6	4.066	0.065	3.905	
Flumequine	42835-25-6	3.566	0.049	3.803	
Oxolinic acid	14698-29-4	3.119	0.046	3.712	
Pyridaphenthion	119-12-0	4.400	0.083	3.973	
1,2,4-trichloro-5-nitrobenzene	89-69-0	3.325	0.046	3.754	
2-chloro-1-fluoro-4-nitrobenzene	350-30-1	0.984	0.163	3.277	1.94
3,4-dichloronitrobenzene	99-54-7	2.118	0.073	3.508	1.61
4-amino-2,6-dinitrotoluene	19406-51-0	1.176	0.143	3.316	
4-chloro-3-methylnitrobenzene	13290-74-9	0.513	0.219	3.181	
1,3-diphenylguanidine	102-06-7	0.371	0.239	3.152	**1.09**
2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	901-44-0	3.365	0.046	3.762	1.18
2,3,4,4'-tetrahydroxybenzophenon	31127-54-5	2.776	0.049	3.642	**0.84**
2,4-diamino-6-phenyl-*s*-triazine	91-76-9	0.214	0.261	3.120	
2-hydroxy-4-methoxybenzophenone	131-57-7	1.814	0.091	3.446	1.78
3,4,4'-trichlorodiphenylurea	101-20-2	4.145	0.069	3.921	
4,4'-diaminodiphenyl ether	101-80-4	0.474	0.225	3.173	
4,4'-dihydroxydiphenylmethane	620-92-8	0.395	0.235	3.157	1.19
4,4'-methylenedianiline	101-77-9	0.312	0.247	3.140	0.98
Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-	51963-82-7	2.589	0.054	3.604	1.08
Benzophenone	119-61-9	0.160	0.269	3.109	
Bis(4-hydroxyphenyl)sulfone	80-09-1	3.095	0.046	3.707	
Bisphenol A	80-05-7	1.961	0.082	3.476	1.46
Dibromocresyl glycidyl ether	30171-80-3	3.546	0.049	3.799	2.39

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
di-*p*-tolylamine	620-93-9	0.312	0.247	3.140	**2.66**
n,*n*'-bis(2-methylphenyl)guanidine	97-39-2	1.622	0.105	3.407	1.10
Procymidone	32809-16-8	3.885	0.058	3.868	
Pyrimethamine	58-14-0	2.398	0.060	3.565	
Sulfadiazine	68-35-9	2.702	0.051	3.627	
Sulfadimethoxine	122-11-2	3.836	0.056	3.858	
Sulfamethazine	57-68-1	3.306	0.046	3.750	
Sulfamethoxazole	723-46-6	2.599	0.054	3.606	
Trimethoprim	738-70-5	3.271	0.046	3.743	
1-benzo[b]thien-2-ylethan-1-one	22720-75-8	0.886	0.174	3.257	1.25
1-chloro-2,4-dinitrobenzene	97-00-7	1.549	0.110	3.392	3.10
2-(1'-cyclohexenyl)cyclohexanone	1502-22-3	0.322	0.245	3.142	0.79
2,2'-dithiobisbenzothiazole	120-78-5	5.043	0.132	4.104	
2,3-dichloro-1,4-napthoquinone	117-80-6	4.125	0.068	3.917	3.86
2-acetoxy-1,4-naphthoquinone	1785-65-5	1.863	0.088	3.456	
2-hydroxy-1,4-naphthoquinone	83-72-7	1.048	0.156	3.290	
2-methyl-1,4-naphthoquinone	58-27-5	0.901	0.172	3.260	
2-methylthio-4-*tert*-butylamino-6-amino-*s*-triazine	30125-65-6	0.827	0.181	3.245	
2-oxabicyclo[2,2,2]octane, 1,3,3-trimethyl-	470-82-6	0.332	0.244	3.144	
2-phenylindole	948-65-2	0.552	0.214	3.189	2.85
3,3'-dichlorobenzidine	91-94-1	3.139	0.046	3.716	2.70
3,4,5-trichlorocatechol	56961-20-7	4.204	0.072	3.933	
3,4,6-trichlorocatechol	32139-72-3	3.733	0.053	3.837	
3,4-dichlorocatechol	3978-67-4	2.569	0.054	3.600	
3,5-dichlorocatechol	13673-92-2	1.318	0.130	3.345	
3,5-di-*tert*-butylsalicylic acid	19715-19-6	2.785	0.049	3.644	1.97

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
3-amino-4-chlorobenzoic acid	2840-28-0	0.611	0.207	3.201	
4,4'-dihydroxy-biphenyl	92-88-6	0.253	0.255	3.128	**1.16**
4,5-dichlorocatechol	3428-24-8	2.064	0.076	3.497	
4,6-dinitro-*o*-cresol	534-52-1	1.303	0.131	3.342	2.26
6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	91-53-2	1.691	0.099	3.421	1.51
6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	2439-01-2	2.206	0.068	3.526	
Benazolin ethyl	25059-80-7	3.296	0.046	3.748	
Betariboacetate	13035-61-5	3.193	0.046	3.727	
Bismerthiazol	79319-85-0	1.544	0.111	3.391	
Buparvaquone	88426-33-9	4.469	0.087	3.987	
Butachlor	23184-66-9	3.473	0.048	3.784	3.05
Butylbenzyl phthalate	85-68-7	3.001	0.046	3.688	
Captopril	62571-86-2	1.647	0.103	3.412	
Cinmethylin	87818-31-3	3.252	0.046	3.739	
Cotinine	486-56-6	0.209	0.261	3.119	
Cyclosulfamuron	136849-15-5	5.195	0.146	4.135	
Diallyl phthalate	131-17-9	1.774	0.093	3.438	2.75
Diethyl phthalate	84-66-2	1.303	0.131	3.342	
Diisobutyl phthalate	84-69-5	1.877	0.087	3.459	1.92
Dinitramine	29091-05-2	3.684	0.052	3.827	
Diphenylpropanediol		2.275	0.065	3.540	
Droperidol	548-73-2	4.518	0.090	3.997	
Enrofloxacin	93106-60-6	4.896	0.119	4.074	
Flumazenil	78755-81-4	4.238	0.074	3.940	
Haloxyfop R	72619-32-0	4.739	0.106	4.042	
Hydrogenatedbisphenol A	80-04-6	2.471	0.058	3.580	1.10

Table A1. (Continued).

NAME	CAS	Pred. $pEC_{50,EROD[PLHC-1]}$ by Eq. 4.1 (mM)	Hat value (*h**=0.273)	ATS3m	pLC_{50} (mM)
Medazepam	2898-12-6	3.369	0.046	3.763	
n-(*tert*-Butyl)-2-benzothiazolylsulfenamide	95-31-8	2.520	0.056	3.590	2.23
n-cyclohexyl-2-benzothiazolylsulfenamide	95-33-0	2.888	0.047	3.665	2.10
Olaquindox	23696-28-8	3.335	0.046	3.756	
Ondansetron	99614-02-5	3.698	0.052	3.830	
o-tolidine	119-93-7	1.495	0.115	3.381	1.21
Phenol,4,4',4''-ethylidynetris-	27955-94-8	3.742	0.054	3.839	
Propyl gallate	121-79-9	1.539	0.111	3.390	
Propyzamide	23950-58-5	2.363	0.062	3.558	
Pyrazosulfuron ethyl	93697-74-6	5.549	0.185	4.207	
Secobarbital	76-73-3	2.879	0.048	3.663	
Simazine	122-34-9	0.361	0.240	3.150	
Sulfaquinoxaline	59-40-5	3.865	0.057	3.864	
tert-butyl 2-ethylperoxyhexanoate	3006-82-4	0.768	0.188	3.233	1.67
Tetrachlorocatechol	1198-55-6	5.686	0.201	4.235	
Tetrachlorohydroquinone	87-87-6	5.495	0.179	4.196	
Tetrahydromethylphthalic anhydride	11070-44-3	0.528	0.217	3.184	
Thiopental	76-75-5	3.090	0.046	3.706	
Triclosan	3380-34-5	3.605	0.050	3.811	2.64
Trifluralin	1582-09-8	3.998	0.062	3.891	
Triphenyl phosphate	115-86-6	3.973	0.061	3.886	2.40
Tris (2-chloroethyl) phosphate	115-96-8	2.496	0.057	3.585	

*Bold values indicate pLC_{50} data with very weak relationship with the $pEC_{50,EROD[PLHC-1]}$ prediction.

APPENDIX B: EXTERNAL SET CHEMICALS FOR $pEC_{50,NR[PLHC-1]}$ MODEL

Table B1. External set chemicals for $pEC_{50,NR[PLHC-1]}$ model.

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
5-fluorocytosine	2022-85-7	-1.215	0.164	0	0	0.4	
5-fluorouracil	51-21-8	-1.215	0.164	0	0	0.4	
Acebutolol	37517-30-9	1.212	0.101	3	0	1	
Acetaminophen	103-90-2	-0.013	0.053	2	0	0.6	
Acridine	260-94-6	-1.007	0.146	0	0	0.5	
Acrivastine	87848-99-5	0.007	0.084	4	1	0.8	
Acyclovir	59277-89-3	0.819	0.110	2	0	1	
Alprenolol	13655-52-2	0.819	0.110	2	0	1	
Amitriptyline	50-48-6	0.980	0.044	4	0	0.7	
Amobarbital	57-43-2	0.034	0.179	0	0	1	
Ampicillin	69-53-4	-0.755	0.136	1	1	1	
Antipyrine	60-80-0	-0.406	0.086	1	0	0.6	
Atovaquone	95233-18-4	1.188	0.057	4	0	0.8	
Bumetanide	28395-03-1	0.400	0.110	5	1	0.8	
Bupropion	34841-39-9	0.819	0.110	2	0	1	
Caffeine	58-08-2	0.403	0.065	2	0	0.8	
Camptothecin	7689-03-4	0.403	0.065	2	0	0.8	
Capecitabine	154361-50-9	0.034	0.179	0	0	1	
Carvedilol	72956-09-3	0.796	0.053	3	0	0.8	
Cefurexime axetil	64544-07-6	0.010	0.094	1	0	0.8	
Cefuroxime	55268-75-2	-1.587	0.108	1	1	0.6	
Cephalexin	15686-71-2	-0.755	0.136	1	1	1	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Chloramphenicol	56-75-7	0.819	0.110	2	0	1	
Chlorotetracycline	57-62-5	1.396	0.079	4	0	0.9	
Chlorpromazine	50-53-3	1.581	0.079	5	0	0.8	
Chlorpropamide	94-20-2	0.403	0.065	2	0	0.8	
Cimetidine	51481-61-9	0.819	0.110	2	0	1	
Ciprofloxacin	85721-33-1	0.423	0.125	4	1	1	
Clonazepam	1622-61-3	0.749	0.069	5	0	0.4	
Clonidine	4205-90-7	0.796	0.053	3	0	0.8	
Cytarabine	147-94-4	-0.175	0.156	0	0	0.9	
Deoxytetracycline	564-25-0	0.796	0.053	3	0	0.8	
Desipramine	50-47-5	1.188	0.057	4	0	0.8	
Doxycycline	564-25-0	0.796	0.053	3	0	0.8	
Estradiol	50-28-2	1.212	0.101	3	0	1	
Ethinylestradiol	57-63-6	1.212	0.101	3	0	1	
Famciclovir	104227-87-4	0.426	0.136	1	0	1	
Famotidine	76824-35-6	0.195	0.055	2	0	0.7	
Finasteride	98319-26-7	0.034	0.179	0	0	1	
Fleroxacin	79660-72-3	0.816	0.154	5	1	1	
Florfenicol	73231-34-2	0.819	0.110	2	0	1	
Flunitrazepam	1622-62-4	0.749	0.069	5	0	0.4	
Flutamide	13311-84-7	0.796	0.053	3	0	0.8	
Furaltadone	139-91-3	0.819	0.110	2	0	1	
Fusidic acid	6990-06-3	-1.356	0.156	0	1	0.9	
Gemcitabine	95058-81-4	0.034	0.179	0	0	1	
Glibenclamid	10238-21-8	1.997	0.134	5	0	1	
Hydrochlorothiazide	58-93-5	0.148	0.074	4	0	0.3	
Hydrocortisone	50-23-7	0.034	0.179	0	0	1	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Imipramine	50-49-7	1.188	0.057	4	0	0.8	
Indomethacin	53-86-1	-0.016	0.099	5	1	0.6	
Isoniazid	54-85-3	-0.406	0.086	1	0	0.6	
Isotretinoin	4759-48-2	-1.148	0.174	0	1	1	
Itraconazole	84625-61-6	1.743	0.142	7	0	0.5	
Ketoconazole	65277-42-1	1.789	0.102	5	0	0.9	
Ketoprofen	22071-15-4	-0.802	0.070	3	1	0.6	
Labetalol	36894-69-6	1.604	0.109	4	0	1	
Lamotrigine	84057-84-1	1.142	0.100	6	0	0.4	
Lansoprazole	103577-45-3	1.188	0.057	4	0	0.8	
Levofloxacin	100986-85-4	0.816	0.154	5	1	1	
Lidocaine	137-58-6	1.212	0.101	3	0	1	
Lomefloxacin	98079-51-7	0.816	0.154	5	1	1	
L-tryptophan	73-22-3	-1.795	0.122	1	1	0.5	
Metaxalone	1665-48-1	1.212	0.101	3	0	1	
Metformin	657-24-9	0.034	0.179	0	0	1	
Methylprednisolone	83-43-2	0.034	0.179	0	0	1	
Metipranolol	22664-55-7	1.997	0.134	5	0	1	
Metoprolol	37350-58-6	0.819	0.110	2	0	1	
Metronidazole	443-48-1	0.403	0.065	2	0	0.8	
Midazolam	59467-70-8	1.951	0.143	7	0	0.6	
Nadolol	42200-33-9	1.212	0.101	3	0	1	
Naproxen	22204-53-1	-0.779	0.081	2	1	0.8	
Nicotine	22083-74-5	0.010	0.094	1	0	0.8	
Nisoldipine	63675-72-9	0.819	0.110	2	0	1	
Nitrofurazone	59-87-0	-0.429	0.075	2	0	0.4	
Norfloxacin	70458-96-7	0.423	0.125	4	1	1	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Novobiocin	303-81-1	2.159	0.154	7	0	0.7	
Ofloxacin	82419-36-1	0.816	0.154	5	1	1	
Oxprenolol	6452-71-7	0.819	0.110	2	0	1	
Oxyphenbutazone	129-20-4	0.796	0.053	3	0	0.8	
Oxytetracycline	79-57-2	0.588	0.041	3	0	0.7	
Paracetamol	103-90-2	-0.013	0.053	2	0	0.6	
Paroxetine	61869-08-7	1.581	0.079	5	0	0.8	
Pentobarbital	76-74-4	0.034	0.179	0	0	1	
Phenobarbital	50-06-6	0.010	0.094	1	0	0.8	
Phenoxymethylpenicillinic Acid	87-08-1	-0.755	0.136	1	1	1	
Phenylbutazone	50-33-9	0.403	0.065	2	0	0.8	
Phenytoin	57-41-0	-0.013	0.053	2	0	0.6	
Pindolol	13523-86-9	0.426	0.136	1	0	1	
Piperazine	110-85-0	-0.383	0.141	0	0	0.8	
Prazosin	19216-56-9	1.581	0.079	5	0	0.8	
Prednisolone	50-24-8	0.034	0.179	0	0	1	
Procaine	59-46-1	0.819	0.110	2	0	1	
Progesterone	57-83-0	0.034	0.179	0	0	1	
Promazine	58-40-2	1.188	0.057	4	0	0.8	
Propylthiouracil	51-52-5	0.034	0.179	0	0	1	
Quinidine	56-54-2	0.403	0.065	2	0	0.8	
Quinine	130-95-0	0.403	0.065	2	0	0.8	
Ranitidine	66357-35-5	0.819	0.110	2	0	1	
Risperidone	106266-06-2	0.403	0.065	2	0	0.8	
Sotalol	3930-20-9	0.819	0.110	2	0	1	
Sulfaguanidine	57-67-0	-0.221	0.060	2	0	0.5	
Sulfaphenazole	526-08-9	0.772	0.039	4	0	0.6	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (h*=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Sulfasalazine	599-79-1	-0.040	0.158	6	1	0.4	
Sumatriptan	103628-46-2	0.403	0.065	2	0	0.8	
Terazosin	63590-64-7	1.604	0.109	4	0	1	
Terbinafine	91161-71-6	-0.198	0.086	1	0	0.7	
Testosterone	58-22-0	0.034	0.179	0	0	1	
Theophylline	58-55-9	0.403	0.065	2	0	0.8	
Timolol	26839-75-8	0.819	0.110	2	0	1	
Tolazamide	1156-19-0	0.819	0.110	2	0	1	
Tolbutamide	64-77-7	0.819	0.110	2	0	1	
Trenbolone acetate	10161-34-9	0.034	0.179	0	0	1	
Triamterene	396-01-0	0.541	0.087	5	0	0.3	
Trifluprom azine	146-54-3	1.581	0.079	5	0	0.8	
Warfarin	81-81-2	0.380	0.038	3	0	0.6	
2,6-dichlorobenzonitrile	1194-65-6	-0.244	0.078	3	0	0.3	
Acetochlor	34256-82-1	1.212	0.101	3	0	1	
Acroleine	107-02-8	-1.215	0.164	0	0	0.4	
Ametryn	834-12-8	1.212	0.101	3	0	1	
Anilofos	64249-01-0	0.819	0.110	2	0	1	
Azocyclotin	41083-11-8	-0.175	0.156	0	0	0.9	
Benalaxyl	71626-11-4	1.604	0.109	4	0	1	
Benomyl	17804-35-2	0.010	0.094	1	0	0.8	
Bensulfuron-methyl	83055-99-6	1.373	0.063	5	0	0.7	
Beta-cyfluthrin	68359-37-5	1.188	0.057	4	0	0.8	
Bromoxynil	1689-84-5	0.356	0.054	4	0	0.4	
Chloridazon	1698-60-8	-0.822	0.111	1	0	0.4	
Chlorimuron-ethyl	90982-32-4	0.749	0.069	5	0	0.4	
Chlorothalonil	1897-45-6	1.142	0.100	6	0	0.4	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (h*=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Chlorotoluron	15545-48-9	1.212	0.101	3	0	1	
Chlorsulfuron	64902-72-3	0.749	0.069	5	0	0.4	
Cyanazine	21725-46-2	1.212	0.101	3	0	1	
Cyhexatin	13121-70-5	-1.007	0.146	0	0	0.5	
Decanoic acid	334-48-5	-1.564	0.147	0	1	0.8	
Diclofop methyl	51338-27-3	1.165	0.057	5	0	0.6	
Diclofop P	40843-25-2	-0.016	0.099	5	1	0.6	
Dimethachlon	24096-53-5	0.380	0.038	3	0	0.6	
Endosulfan	115-29-7	-0.799	0.136	0	0	0.6	
Ethalfluraline	55283-68-6	1.188	0.057	4	0	0.8	
Ethametsulfuron	111353-84-5	-0.640	0.145	5	1	0.3	
Ethoxyquin	91-53-2	1.212	0.101	3	0	1	1.51
Fenamiphos	22224-92-6	1.212	0.101	3	0	1	
Fenhexamid	126833-17-8	1.604	0.109	4	0	1	
Fenoxaprop	95617-09-7	-0.825	0.102	4	1	0.4	
Fenpropidin	67306-00-7	0.611	0.084	2	0	0.9	
Fentin hydroxide	76-87-9	0.172	0.043	3	0	0.5	
Fluazifop P	83066-88-0	-0.825	0.102	4	1	0.4	
Flumetsulam	98967-40-9	0.749	0.069	5	0	0.4	
Fluroxypyr	69377-81-7	-0.432	0.122	5	1	0.4	
Iprodione	36734-19-7	1.212	0.101	3	0	1	
Cybutryne	28159-98-0	1.212	0.101	3	0	1	
Isoproturon	34123-59-6	0.819	0.110	2	0	1	
Lindane	58-89-9	-0.591	0.134	0	0	0.7	
Mefenacet	73250-68-7	-0.013	0.053	2	0	0.6	
Metalaxyl	57837-19-1	1.212	0.101	3	0	1	
Methabenzthiazuron	18691-97-9	-0.406	0.086	1	0	0.6	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Metolachlor	51218-45-2	1.212	0.101	3	0	1	
Metribuzin	21087-64-9	0.034	0.179	0	0	1	
Metsulfuron-methyl	74223-64-6	0.541	0.087	5	0	0.3	
Molinate	2212-67-1	0.034	0.179	0	0	1	
Nicosulfuron	111991-09-4	1.373	0.063	5	0	0.7	
Oxadiargyl	39807-15-3	1.188	0.057	4	0	0.8	
Oxadiazon	19666-30-9	1.604	0.109	4	0	1	
Oxadixyl	77732-09-3	1.212	0.101	3	0	1	
Pendimethalin	40487-42-1	1.997	0.134	5	0	1	
Phenmedipham	13684-63-4	0.772	0.039	4	0	0.6	
Pretilachlor	51218-49-6	1.212	0.101	3	0	1	2.11
Prometryn	7287-19-6	1.212	0.101	3	0	1	
Quinclorac	84087-01-4	-1.218	0.099	3	1	0.4	
Quizalofop P	94051-08-8	-0.825	0.102	4	1	0.4	
Tebuthiuron	34014-18-1	0.819	0.110	2	0	1	
Terbumeton	33693-04-8	1.212	0.101	3	0	1	
Terbuthylazine	5915-41-3	1.212	0.101	3	0	1	
Terbutryn	886-50-0	1.212	0.101	3	0	1	
Thiobencarb	28249-77-6	0.819	0.110	2	0	1	2.30
Thiophanate methyl	23564-05-8	0.403	0.065	2	0	0.8	
Tribenuron	106040-48-6	-0.640	0.145	5	1	0.3	
1,1,2,2-tetrachloroethane	79-34-5	-0.591	0.134	0	0	0.7	
1,1,2-trichloroethane	79-00-5	-0.591	0.134	0	0	0.7	
1,1-dichloroethane	75-34-3	-1.007	0.146	0	0	0.5	
1,1-dichloroethylene	75-35-4	-1.007	0.146	0	0	0.5	0.33
1,2,3-trichloropropane	96-18-4	-1.007	0.146	0	0	0.5	
1,2-dibromo-3-chloropropane	96-12-8	-1.007	0.146	0	0	0.5	0.78

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
1,2-dichloroethane	107-06-2	-1.007	0.146	0	0	0.5	
1,2-dichloropropane	78-87-5	-1.007	0.146	0	0	0.5	-0.15
1,3-dibromopropane	109-64-8	-1.007	0.146	0	0	0.5	1.41
1,3-dichloropropane	142-28-9	-1.007	0.146	0	0	0.5	
1,3-dichloropropene	542-75-6	-0.591	0.134	0	0	0.7	1.87
1-chlorobutane	109-69-3	-1.007	0.146	0	0	0.5	
3,4-dichlorobut-1-ene	760-23-6	-0.175	0.156	0	0	0.9	0.67
Acetic acid, bromo-, 2-butene-1,4-diyl ester	20679-58-7	0.034	0.179	0	0	1	
Bromodichloromethane	75-27-4	-1.007	0.146	0	0	0.5	0.77
Carbon tetrachloride	56-23-5	-1.423	0.191	0	0	0.3	**1.31**
Cyclohexane	110-82-7	-1.007	0.146	0	0	0.5	
Dibromochloromethane	124-48-1	-1.007	0.146	0	0	0.5	0.42
Ethyl bromide	74-96-4	-1.007	0.146	0	0	0.5	
Ethylcyclohexane	1678-91-7	-1.007	0.146	0	0	0.5	**2.18**
Methylcyclohexane	108-87-2	-1.007	0.146	0	0	0.5	**1.67**
Pentachloroethane	76-01-7	-1.007	0.146	0	0	0.5	
Tetrachloroethene	127-18-4	-1.423	0.191	0	0	0.3	1.07
Trichloroethylene	79-01-6	-1.423	0.191	0	0	0.3	0.54
1-(*n*-phenylamino)-naphthalene	90-30-2	-0.221	0.060	2	0	0.5	2.50
1,2-dimethylnaphthalene	573-98-8	0.195	0.055	2	0	0.7	1.83
1,3-dimethylnapthalene	575-41-7	0.195	0.055	2	0	0.7	1.94
1,5-napthalenediamine	2243-62-1	-0.013	0.053	2	0	0.6	0.97
1,8-naphthylenediamine	479-27-6	-0.013	0.053	2	0	0.6	1.45
1-methylnaphthalene	90-12-0	-0.614	0.095	1	0	0.5	1.40
2,7-dimethylnaphthalene	582-16-1	0.195	0.055	2	0	0.7	1.94
2-methylnaphthalene	91-57-6	-0.614	0.095	1	0	0.5	1.87
3-hydroxy-2-naphthoic acid	92-70-6	-1.611	0.113	2	1	0.4	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Isopropylnaphthalene	29253-36-9	-0.198	0.086	1	0	0.7	2.36
n-phenyl-2-naphthylamine	135-88-6	-0.221	0.060	2	0	0.5	
ß-naphthol	135-19-3	-0.614	0.095	1	0	0.5	1.56
ß-naphthylamine	91-59-8	-0.614	0.095	1	0	0.5	1.56
Tetralin	119-64-2	0.195	0.055	2	0	0.7	1.34
1,2,3-trichlorobenzene	87-61-6	-0.244	0.078	3	0	0.3	1.75
1,2-dichlorobenzene	95-50-1	-0.221	0.060	2	0	0.5	1.59
1,3-dichlorobenzene	541-73-1	-0.221	0.060	2	0	0.5	1.41
1,4-benzenediamine, *n*,*n*'-bis(1-methylpropyl)-	101-96-2	0.819	0.110	2	0	1	**2.78**
1,4-dichlorobenzene	106-46-7	-0.221	0.060	2	0	0.5	**1.82**
2,4,6-trichlorophenylhydrazine	5329-12-4	0.772	0.039	4	0	0.6	2.33
2,4,6-trinitrotoluene	118-96-7	0.356	0.054	4	0	0.4	
2,4-diamino-6-nitrotoluene	6629-29-4	0.772	0.039	4	0	0.6	
2,4-diaminotoluene	95-80-7	0.380	0.038	3	0	0.6	
2,4-dichlorotoluene	95-73-8	0.172	0.043	3	0	0.5	
2,4-dinitrotoluene	121-14-2	-0.036	0.056	3	0	0.4	
2,5-diaminotoluene	95-70-5	0.380	0.038	3	0	0.6	2.81
2,5-dichlorotoluene	19398-61-9	0.172	0.043	3	0	0.5	1.60
2,6-diamino-4-nitrotoluene	59229-75-3	0.772	0.039	4	0	0.6	
2,6-diaminotoluene	823-40-5	0.380	0.038	3	0	0.6	
2,6-dichlorotoluene	118-69-4	0.172	0.043	3	0	0.5	1.85
2,6-dinitrotoluene	606-20-2	-0.036	0.056	3	0	0.4	0.73
2-amino-4,6-dinitrotoluene	35572-78-2	0.356	0.054	4	0	0.4	
2-amino-4-nitrotoluene	99-55-8	0.380	0.038	3	0	0.6	
2-amino-6-nitrotoluene	603-83-8	0.380	0.038	3	0	0.6	
2-chlorohydroquinonedimethylether	2100-42-7	0.796	0.053	3	0	0.8	0.79
2-chlorotoluene	95-49-8	-0.221	0.060	2	0	0.5	1.22

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
2-methyl-4-chlorophenoxyacetic acid	94-74-6	-0.386	0.074	3	1	0.8	
2-phenylpropene	98-83-9	-0.614	0.095	1	0	0.5	1.21
3,4-dichlorotoluene	95-75-0	0.172	0.043	3	0	0.5	1.57
3,5-bis(trifluoromethyl)benzylamine	85068-29-7	0.588	0.041	3	0	0.7	1.34
3-chlorotoluene	108-41-8	-0.221	0.060	2	0	0.5	
4,6-dichlororesorcinol	137-19-9	0.356	0.054	4	0	0.4	
4-allyl-1,2-dimethoxybenzene	93-15-2	1.212	0.101	3	0	1	1.10
4-chlororesorcinol	95-88-5	0.380	0.038	3	0	0.6	
4-chlorotoluene	106-43-4	-0.221	0.060	2	0	0.5	1.32
4-nitrotoluene-2-sulphonic acid	121-03-9	-1.218	0.099	3	1	0.4	
4-*tert*-butyltoluene	98-51-1	0.611	0.084	2	0	0.9	1.94
4-toluenesulfonyl chloride	98-59-9	-0.013	0.053	2	0	0.6	0.39
Atenolol	29122-68-7	0.819	0.110	2	0	1	
Benzalacetone	122-57-6	-0.614	0.095	1	0	0.5	1.41
Benzotrifluoride	98-08-8	-0.614	0.095	1	0	0.5	1.01
Benzyl alcohol	100-51-6	-0.614	0.095	1	0	0.5	
Benzyl cyanide	140-29-4	-0.614	0.095	1	0	0.5	
Benzylamine	100-46-9	-0.614	0.095	1	0	0.5	
Bromobenzene	108-86-1	-0.614	0.095	1	0	0.5	1.56
Butylbenzene	104-51-8	-0.198	0.086	1	0	0.7	1.61
Catechol	120-80-9	-0.013	0.053	2	0	0.6	
Chlorobenzene	108-90-7	-0.614	0.095	1	0	0.5	1.23
Cyclohexylbenzene	827-52-1	-0.198	0.086	1	0	0.7	2.13
Diisopropylbenzene	25321-09-9	0.611	0.084	2	0	0.9	2.30
Dimethyl phthalate	131-11-3	-0.013	0.053	2	0	0.6	
Diuron	330-54-1	0.796	0.053	3	0	0.8	
Divinylbenzene	1321-74-0	-0.221	0.060	2	0	0.5	1.49

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Ethylbenzene	100-41-4	-0.198	0.086	1	0	0.7	
Fenobucarb	3766-81-2	0.819	0.110	2	0	1	1.33
Isopropylbenzene	98-82-8	-0.198	0.086	1	0	0.7	
Metaxylene hexafluoride	402-31-3	-0.221	0.060	2	0	0.5	1.47
m-phenylenebis(methylamine)	1477-55-0	0.403	0.065	2	0	0.8	0.19
m-phenylenediamine	108-45-2	-0.013	0.053	2	0	0.6	
n-propylbenzene	103-65-1	-0.198	0.086	1	0	0.7	
o-phenylenediamine	95-54-5	-0.013	0.053	2	0	0.6	1.37
o-toluenesulfonamide	88-19-7	-0.013	0.053	2	0	0.6	
Oxyfluorfen	42874-03-3	1.142	0.100	6	0	0.4	
p-cymene	99-87-6	0.195	0.055	2	0	0.7	1.83
p-phenylenediamine	106-50-3	-0.013	0.053	2	0	0.6	**3.21**
Propoxur	114-26-1	0.819	0.110	2	0	1	
Resorcinol	108-46-3	-0.013	0.053	2	0	0.6	
Sulphanilamide	63-74-1	-0.013	0.053	2	0	0.6	
Thiamphenicol	15318-45-3	0.819	0.110	2	0	1	
Toluene	108-88-3	-0.614	0.095	1	0	0.5	0.57
Trans-cinnamic acid	140-10-3	-1.587	0.108	1	1	0.6	
2-(1,1-dimethyl)-4,6-dimethylphenol	1879-09-0	1.396	0.079	4	0	0.9	1.85
2,3,4,5-tetrachlorophenol	4901-51-3	0.541	0.087	5	0	0.3	
2,3,5,6-tetrachlorophenol	935-95-5	0.541	0.087	5	0	0.3	
2,3,5-trichlorophenol	933-78-8	0.148	0.074	4	0	0.3	
2,3,5-trimethylphenol	697-82-5	0.980	0.044	4	0	0.7	
2,3,6-trichlorophenol	933-75-5	0.148	0.074	4	0	0.3	
2,3,6-trimethylphenol	2416-94-6	0.980	0.044	4	0	0.7	
2,3-dimethylphenol	526-75-0	0.588	0.041	3	0	0.7	
2,4,6-tribromophenol	118-79-6	0.148	0.074	4	0	0.3	2.34

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
2,4,6-trimethylphenol	527-60-6	0.980	0.044	4	0	0.7	
2,4-dibromophenol	615-58-7	-0.244	0.078	3	0	0.3	1.84
2,4-di-*tert*-butylphenol	96-76-4	1.004	0.073	3	0	0.9	2.48
2,5-dimethylphenol	95-87-4	0.588	0.041	3	0	0.7	1.33
2,6-dimethylphenol	576-26-1	0.588	0.041	3	0	0.7	0.91
2,6-di-*sec*-butylphenol	5510-99-6	1.004	0.073	3	0	0.9	**3.14**
2-allylphenol	1745-81-9	-0.221	0.060	2	0	0.5	
2-ethylphenol	90-00-6	0.195	0.055	2	0	0.7	
2-methoxyphenol	90-05-1	-0.013	0.053	2	0	0.6	
2-*n*-propylphenol	644-35-9	0.195	0.055	2	0	0.7	
2-*tert*-butyl phenol	88-18-6	0.611	0.084	2	0	0.9	1.61
2-*tert*-butyl-*p*-cresol	2409-55-4	1.004	0.073	3	0	0.9	1.94
3,4,5-trichloroguaiacol	57057-83-7	0.749	0.069	5	0	0.4	
3,4,5-trichlorophenol	609-19-8	0.148	0.074	4	0	0.3	
3,4-dimethylphenol	95-65-8	0.588	0.041	3	0	0.7	
3-ethylphenol	620-17-7	0.195	0.055	2	0	0.7	
3-trifluoromethyl-4-nitrophenol	88-30-2	-0.036	0.056	3	0	0.4	
4-(1-methylethenyl)phenol	4286-23-1	-0.221	0.060	2	0	0.5	1.16
4-(2,4-dichlorophenoxy)-phenol	40843-73-0	0.749	0.069	5	0	0.4	
4,5,6-trichloroguaiacol	2668-24-8	0.749	0.069	5	0	0.4	
4,5-dichloroguaiacol	2460-49-3	0.772	0.039	4	0	0.6	
4-chloro-2-methylphenol	1570-64-5	0.172	0.043	3	0	0.5	
4-chloro-2-nitrophenol	89-64-5	-0.036	0.056	3	0	0.4	1.20
4-ethylphenol	123-07-9	0.195	0.055	2	0	0.7	
4-*n*-nonylphenol	104-40-5	0.403	0.065	2	0	0.8	
4-*n*-octylphenol	1806-26-4	0.611	0.084	2	0	0.9	**3.37**
4-pentylphenol	14938-35-3	0.611	0.084	2	0	0.9	2.07

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
6-*tert*-butyl-*m*-cresol	88-60-8	1.004	0.073	3	0	0.9	1.78
6-*tert*-butyl-*o*-cresol	2219-82-1	1.004	0.073	3	0	0.9	1.58
Butylated hydroxyanisole	25013-16-5	1.212	0.101	3	0	1	1.49
Methyl *p*-hydroxybenzoate	99-76-3	-0.013	0.053	2	0	0.6	0.40
o-*sec*-butylphenol	89-72-5	0.611	0.084	2	0	0.9	1.40
p-*sec*-butylphenol	99-71-8	0.611	0.084	2	0	0.9	1.76
p-*tert*-butylphenol	98-54-4	0.611	0.084	2	0	0.9	
Tetrachloroguaiacol	2539-17-5	1.142	0.100	6	0	0.4	
Thymol	89-83-8	1.004	0.073	3	0	0.9	1.50
Trichlorosyringol	2539-26-6	2.390	0.175	6	0	1	
1-nitropyrene	5522-43-0	-0.822	0.111	1	0	0.4	
2-acetamidophenoxazin-3-one	1916-55-8	-0.429	0.075	2	0	0.4	
2-amino-7-methoxyphenoxazin-3-one		-0.036	0.056	3	0	0.4	
2-aminophenoxazin-3-one	1916-59-2	-0.429	0.075	2	0	0.4	
9-vinylcarbazole	1484-13-5	-1.007	0.146	0	0	0.5	
Acenaphthene	83-32-9	0.195	0.055	2	0	0.7	
Dibenzo[b,f]cyclohepten-1-one	2222-33-5	0.564	0.042	4	0	0.5	
Dibenzothiophene	132-65-0	-1.007	0.146	0	0	0.5	**2.12**
Fluorene	86-73-7	0.564	0.042	4	0	0.5	
Phenanthrene	85-01-8	-1.007	0.146	0	0	0.5	**2.10**
Phenothiazine	92-84-2	0.564	0.042	4	0	0.5	2.41
1-decanol	112-30-1	-1.423	0.191	0	0	0.3	**1.75**
1-nonanol	143-08-8	-1.423	0.191	0	0	0.3	**1.65**
2-(2-butoxyethoxy)ethanol	112-34-5	-0.799	0.136	0	0	0.6	
2-butoxyethanol	111-76-2	-0.799	0.136	0	0	0.6	
2-isoproxyethanol	109-59-1	-0.799	0.136	0	0	0.6	
3-pentanol	71-41-0	-1.423	0.191	0	0	0.3	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Cyclohexanol	108-93-0	-1.007	0.146	0	0	0.5	
Hexanol	111-27-3	-1.423	0.191	0	0	0.3	
Isodecyl alcohol	25339-17-7	-1.423	0.191	0	0	0.3	**1.43**
1,6-hexanediamine	124-09-4	-0.799	0.136	0	0	0.6	**0.21**
2-(dibutylamino)ethanol	102-81-8	-1.215	0.164	0	0	0.4	**0.78**
2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)	6864-37-5	-0.799	0.136	0	0	0.6	**1.04**
2-amino-2-ethylpropanediol	115-70-8	-1.007	0.146	0	0	0.5	
Cyclohexylamine	108-91-8	-1.007	0.146	0	0	0.5	**0.48**
Dibutylamine	111-92-2	-1.423	0.191	0	0	0.3	
Diethanolamine	111-42-2	-0.591	0.134	0	0	0.7	
Diethylamine	109-89-7	-1.215	0.164	0	0	0.4	0.43
Diethylnitrosamine	55-18-5	0.034	0.179	0	0	1	
Diisopropylamine	108-18-9	-1.423	0.191	0	0	0.3	
Dimethylamine	124-40-3	-0.799	0.136	0	0	0.6	
n,n,n',n'-tetramethylhexamethylenediamine	111-18-2	-1.215	0.164	0	0	0.4	
Piperidine	110-89-4	-1.007	0.146	0	0	0.5	
t-butylamine	75-64-9	-1.215	0.164	0	0	0.4	
Triethylamine	121-44-8	-1.423	0.191	0	0	0.3	0.63
2-(dimethylamino)ethyl methacrylate	2867-47-2	0.034	0.179	0	0	1	0.92
2-ethylhexyl methacrylate	688-84-6	0.034	0.179	0	0	1	1.85
2-hydroxyethyl acrylate	818-61-1	0.034	0.179	0	0	1	1.25
2-hydroxyethyl methacrylate	868-77-9	0.034	0.179	0	0	1	
Ethyl trichloroacetate	515-84-4	0.034	0.179	0	0	1	0.64
Ethylacrylate	140-88-5	0.034	0.179	0	0	1	1.92
Isobutyl acetate	110-19-0	0.034	0.179	0	0	1	0.83
Methyl acrylate	96-33-3	0.034	0.179	0	0	1	1.79
Methyl methacrylate	80-62-6	0.034	0.179	0	0	1	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
n-butyl acrylate	141-32-2	0.034	0.179	0	0	1	1.73
n-butyl methacrylate	97-88-1	0.034	0.179	0	0	1	1.40
Testosterone propionate	57-85-2	0.034	0.179	0	0	1	
Tetracaine	94-24-6	0.819	0.110	2	0	1	
Vinyl acetate	108-05-4	0.034	0.179	0	0	1	1.55
2,3,4-trichloroaniline	634-67-3	0.148	0.074	4	0	0.3	
2,3-dichloroaniline	608-27-5	0.172	0.043	3	0	0.5	
2,3-dimethylaniline	87-59-2	0.588	0.041	3	0	0.7	
2,4,5-trichloroaniline	636-30-6	0.148	0.074	4	0	0.3	
2,4,6-trichloroaniline	634-93-5	0.148	0.074	4	0	0.3	1.57
2,4,6-trimethylaniline	88-05-1	0.980	0.044	4	0	0.7	**0.39**
2,4-dichloroaniline	554-00-7	0.172	0.043	3	0	0.5	1.30
2,4-dimethylaniline	95-68-1	0.588	0.041	3	0	0.7	
2,5-dichloroaniline	95-82-9	0.172	0.043	3	0	0.5	1.87
2,5-dimethylaniline	95-78-3	0.588	0.041	3	0	0.7	
2,6-dichloroaniline	608-31-1	0.172	0.043	3	0	0.5	
2,6-diethylaniline	579-66-8	1.004	0.073	3	0	0.9	
2,6-dimethylaniline	87-62-7	0.588	0.041	3	0	0.7	
2-chloro-5-methylaniline	95-81-8	0.172	0.043	3	0	0.5	1.11
2-chloroaniline	95-51-2	-0.221	0.060	2	0	0.5	1.24
2-ethylaniline	578-54-1	0.195	0.055	2	0	0.7	
2-methyl-4-nitroaniline	99-52-5	0.380	0.038	3	0	0.6	
2-methylaniline	95-53-4	-0.221	0.060	2	0	0.5	-0.15
2-nitroaniline	88-74-4	-0.429	0.075	2	0	0.4	0.52
2-nitro-*p*-anisidine	96-96-8	0.380	0.038	3	0	0.6	0.61
3,4,5-trichloroaniline	634-91-3	0.148	0.074	4	0	0.3	
3,4-dichloroaniline	95-76-1	0.172	0.043	3	0	0.5	1.17

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
3,4-dimethylaniline	95-64-7	0.588	0.041	3	0	0.7	
3,5-dichloroaniline	626-43-7	0.172	0.043	3	0	0.5	
3,5-dimethylaniline	108-69-0	0.588	0.041	3	0	0.7	0.55
3-chloroaniline	108-42-9	-0.221	0.060	2	0	0.5	1.16
3-ethylaniline	587-02-0	0.195	0.055	2	0	0.7	
3-methylaniline	108-44-1	-0.221	0.060	2	0	0.5	
3-nitroaniline	99-09-2	-0.429	0.075	2	0	0.4	0.19
4-chloro-2-nitroaniline	89-63-4	-0.036	0.056	3	0	0.4	1.01
4-chloroaniline	106-47-8	-0.221	0.060	2	0	0.5	1.34
4-ethylaniline	589-16-2	0.195	0.055	2	0	0.7	
4-fluoroaniline	371-40-4	-0.221	0.060	2	0	0.5	
4-isopropylaniline	99-88-7	0.195	0.055	2	0	0.7	0.47
4-methylaniline	106-49-0	-0.221	0.060	2	0	0.5	-0.05
4-nitroaniline	100-01-6	-0.429	0.075	2	0	0.4	0.21
Aniline, *p*-(phenylazo)	60-09-3	0.380	0.038	3	0	0.6	2.75
Chlorfluoroaniline	21397-08-0	0.172	0.043	3	0	0.5	
Dithiodianiline	722-27-0	0.772	0.039	4	0	0.6	
n,*n*-diethylaniline	91-66-7	0.218	0.111	1	0	0.9	0.76
n,*n*'-dimethylaniline	-0.198	0.086	1	0	0.7		
n,*n*-dimethylaniline	121-69-7	-0.198	0.086	1	0	0.7	
n-ethylaniline	103-69-5	-0.198	0.086	1	0	0.7	0.22
p-anisidine	104-94-9	-0.013	0.053	2	0	0.6	
2-ethyl butanoic acid	88-09-5	-1.148	0.174	0	1	1	0.14
2-methyl butanoic acid	600-07-7	-1.148	0.174	0	1	1	
3-methyl butanoic acid	503-74-2	-1.148	0.174	0	1	1	
Acrylic acid	79-10-7	-2.188	0.168	0	1	0.5	0.07
A-fluoro-b-alanine	3821-81-6	-1.148	0.174	0	1	1	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Chloroacetic acid	79-11-8	-1.148	0.174	0	1	1	0.12
Dansylglycine	1091-85-6	-1.195	0.080	2	1	0.6	
Flumequine	42835-25-6	0.007	0.084	4	1	0.8	
Gentisic acid	490-79-9	-1.218	0.099	3	1	0.4	
Heptanoic acid	111-14-8	-1.148	0.174	0	1	1	0.24
Isocyanuric acid	108-80-5	-0.036	0.056	3	0	0.4	
Malonic acid diethylester	105-53-3	0.034	0.179	0	0	1	0.61
Methacrylic acid	79-41-4	-1.148	0.174	0	1	1	
Octanoic acid	124-07-2	-1.564	0.147	0	1	0.8	0.45
Orthoformic acid trimethylester	149-73-5	-0.383	0.141	0	0	0.8	
Oxolinic acid	14698-29-4	-0.409	0.076	4	1	0.6	
Perfluorooctanoic acid	335-67-1	-1.564	0.147	0	1	0.8	
Pivalic acid	75-98-9	-1.148	0.174	0	1	1	
Pyridaphenthion	119-12-0	0.426	0.136	1	0	1	
Sorbic acid	110-44-1	-1.564	0.147	0	1	0.8	0.17
2-methylvaleraldehyde	123-15-9	-0.591	0.134	0	0	0.7	
Capronaldehyde	66-25-1	-0.591	0.134	0	0	0.7	
Crotonaldehyde	4170-30-3	-0.799	0.136	0	0	0.6	2.99
Dodecanal	112-54-9	-1.007	0.146	0	0	0.5	
Glutaraldehyde	111-30-8	0.034	0.179	0	0	1	1.06
Propionaldehyde	123-38-6	-0.383	0.141	0	0	0.8	
1,2,4-trichloro-5-nitrobenzene	89-69-0	0.356	0.054	4	0	0.4	
2-chloro-1-fluoro-4-nitrobenzene	350-30-1	-0.036	0.056	3	0	0.4	1.94
2-chloronitrobenzene	88-73-3	-0.429	0.075	2	0	0.4	
2-nitroanisole	91-23-6	-0.013	0.053	2	0	0.6	0.51
2-nitrotoluene	88-72-2	-0.013	0.053	2	0	0.6	
3,4-dichloronitrobenzene	99-54-7	-0.036	0.056	3	0	0.4	1.61

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
3-chloronitrobenzene	121-73-3	-0.429	0.075	2	0	0.4	
3-nitroanisole	555-03-3	-0.013	0.053	2	0	0.6	0.41
4-amino-2,6-dinitrotoluene	19406-51-0	0.356	0.054	4	0	0.4	
4-amino-2-nitrotoluene	119-32-4	0.380	0.038	3	0	0.6	
4-chloro-3-methylnitrobenzene	13290-74-9	-0.036	0.056	3	0	0.4	
4-chloronitrobenzene	100-00-5	-0.429	0.075	2	0	0.4	
4-methylnitrobenzene	99-99-0	-0.013	0.053	2	0	0.6	0.57
alpha-chloro-4-nitrotoluene	100-14-1	-0.429	0.075	2	0	0.4	**2.45**
Nitrobenzene	98-95-3	-0.822	0.111	1	0	0.4	
1,3-diphenylguanidine	102-06-7	-0.013	0.053	2	0	0.6	1.09
2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	901-44-0	1.604	0.109	4	0	1	1.18
2,3,4,4'-tetrahydroxybenzophenon	31127-54-5	1.142	0.100	6	0	0.4	0.84
2,4-diamino-6-phenyl-*s*-triazine	91-76-9	0.772	0.039	4	0	0.6	
2-hydroxy-4-methoxybenzophenone	131-57-7	0.772	0.039	4	0	0.6	1.78
3,4,4'-trichlorodiphenylurea	101-20-2	0.749	0.069	5	0	0.4	
4,4'-diaminodiphenyl ether	101-80-4	0.772	0.039	4	0	0.6	
4,4'-dihydroxydiphenylmethane	620-92-8	0.772	0.039	4	0	0.6	1.19
4,4'-methylenedianiline	101-77-9	0.772	0.039	4	0	0.6	0.98
Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-	51963-82-7	1.604	0.109	4	0	1	1.08
Benzophenone	119-61-9	-0.221	0.060	2	0	0.5	
Bis(4-hydroxyphenyl)sulfone	80-09-1	0.356	0.054	4	0	0.4	
Bisphenol A	80-05-7	1.188	0.057	4	0	0.8	1.46
Dibenzyl ether	103-50-4	-0.221	0.060	2	0	0.5	1.46
Dibromocresyl glycidyl ether	30171-80-3	0.819	0.110	2	0	1	2.39
Diphenyl ether	101-84-8	-0.221	0.060	2	0	0.5	1.98
Diphenylamine	122-39-4	-0.221	0.060	2	0	0.5	1.41
di-*p*-tolylamine	620-93-9	0.564	0.042	4	0	0.5	2.66

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Hydrazobenzene	122-66-7	-0.013	0.053	2	0	0.6	**3.63**
n,*n*'-bis(2-methylphenyl)guanidine	97-39-2	0.772	0.039	4	0	0.6	1.10
Procymidone	32809-16-8	0.796	0.053	3	0	0.8	
Pyrimethamine	58-14-0	1.558	0.092	6	0	0.6	
Styrene-7,8-oxide	96-09-3	-0.198	0.086	1	0	0.7	1.14
Sulfadiazine	68-35-9	-0.036	0.056	3	0	0.4	
Sulfadimethoxine	122-11-2	1.165	0.057	5	0	0.6	
Sulfamethazine	57-68-1	1.165	0.057	5	0	0.6	
Sulfamethoxazole	723-46-6	0.772	0.039	4	0	0.6	
Tetrabromobisphenol A	79-94-7	2.343	0.212	8	0	0.6	1.77
1,1'-oxybis-butane	142-96-1	-1.423	0.191	0	0	0.3	
1,2,3-trihydroxybenzene	87-66-1	0.380	0.038	3	0	0.6	
1,5-cyclooctadiene	111-78-4	-0.175	0.156	0	0	0.9	0.92
1-benzo[b]thien-2-ylethan-1-one	22720-75-8	-0.614	0.095	1	0	0.5	1.25
1-chloro-2,4-dinitrobenzene	97-00-7	-0.036	0.056	3	0	0.4	3.10
1-cyclohexene-1-carbonitrile	1855-63-6	-0.591	0.134	0	0	0.7	0.42
1-mercaptooctane	111-88-6	-1.423	0.191	0	0	0.3	2.65
1-methoxy-2-propanol	107-98-2	-0.799	0.136	0	0	0.6	
2-(1'-cyclohexenyl)cyclohexanone	1502-22-3	-0.175	0.156	0	0	0.9	0.79
2,2,5,5,-tetramethylhydrofuran	15045-43-9	-1.007	0.146	0	0	0.5	
2,2,6,6-tetramethylpiperidin-4-ol	2403-88-5	-0.799	0.136	0	0	0.6	
2,2'-dithiobisbenzothiazole	120-78-5	-0.429	0.075	2	0	0.4	
2,3,3,3,2',3',3',3'-octachlorodipropyl ether	127-90-2	-1.007	0.146	0	0	0.5	2.81
2,3-dichloro-1,4-napthoquinone	117-80-6	-0.429	0.075	2	0	0.4	3.86
2,4,6-trimethylbenzaldehyde	487-68-3	0.980	0.044	4	0	0.7	1.09
2-acetoxy-1,4-naphthoquinone	1785-65-5	-0.429	0.075	2	0	0.4	
2-aminophenol	95-55-6	-0.013	0.053	2	0	0.6	2.21

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
2-aminopyridine	504-29-0	-0.406	0.086	1	0	0.6	0.93
2-butanone oxime	96-29-7	0.034	0.179	0	0	1	
2-butenedinitrile, (E)-	764-42-1	-1.215	0.164	0	0	0.4	2.34
2-Chlorobenzyl chloride	611-19-8	-0.221	0.060	2	0	0.5	**2.78**
2-decanone	693-54-9	-1.007	0.146	0	0	0.5	
2-hydroxy-1,4-naphthoquinone	83-72-7	-0.429	0.075	2	0	0.4	
2-mercaptobenzothiazole	149-30-4	-0.637	0.099	2	0	0.3	
2-mercaptoethanol	60-24-2	-0.175	0.156	0	0	0.9	**0.43**
2-methyl-1,4-naphthoquinone	58-27-5	-0.013	0.053	2	0	0.6	
2-methyl-4-nitroimidazole	696-23-1	-0.429	0.075	2	0	0.4	
2-methylthio-4-*tert*-butylamino-6-amino-*s*-triazine	30125-65-6	1.212	0.101	3	0	1	
2-oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	470-82-6	-1.007	0.146	0	0	0.5	
2-phenylindole	948-65-2	-0.221	0.060	2	0	0.5	**2.85**
2-phenylphenol	90-43-7	0.172	0.043	3	0	0.5	
2-propenenitrile, 2-chloro-	920-37-6	-1.423	0.191	0	0	0.3	**3.02**
2-propenenitrile	107-13-1	-1.215	0.164	0	0	0.4	**1.02**
2-undecanone	112-12-9	-1.007	0.146	0	0	0.5	
2-vinylpyridine	100-69-6	-0.614	0.095	1	0	0.5	**1.21**
3-(methylthio)propionaldehyde	3268-49-3	-0.175	0.156	0	0	0.9	**1.17**
3,3'-dichlorobenzidine	91-94-1	1.558	0.092	6	0	0.6	**2.70**
3,4,5-trichlorocatechol	56961-20-7	0.749	0.069	5	0	0.4	
3,4,6-trichlorocatechol	32139-72-3	0.749	0.069	5	0	0.4	
3,4-dichlorocatechol	3978-67-4	0.356	0.054	4	0	0.4	
3,5,5-trimethyl-2-cyclohexen-1-one	78-59-1	-0.175	0.156	0	0	0.9	
3,5-dichlorocatechol	13673-92-2	0.356	0.054	4	0	0.4	
3,5-di-*tert*-butylsalicylic acid	19715-19-6	0.423	0.125	4	1	1	**1.97**
3a,4,7,7a-tetrahydro-1H-indene	3048-65-5	-0.591	0.134	0	0	0.7	**1.44**

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
3-amino-1,2,4-triazole	61-82-5	-0.406	0.086	1	0	0.6	
3-amino-2-Butenenitrile	1118-61-2	-0.383	0.141	0	0	0.8	0.68
3-amino-4-chlorobenzoic acid	2840-28-0	-1.218	0.099	3	1	0.4	
3-aminophenol	591-27-5	-0.013	0.053	2	0	0.6	-0.04
3-aminopyridine	462-08-8	-0.406	0.086	1	0	0.6	1.04
4,4,4-trifluorocrotonitrile	406-86-0	-1.007	0.146	0	0	0.5	2.78
4,4'-dihydroxy-biphenyl	92-88-6	0.772	0.039	4	0	0.6	1.16
4,5-dichlorocatechol	3428-24-8	0.356	0.054	4	0	0.4	
4,6-dinitro-*o*-cresol	534-52-1	0.356	0.054	4	0	0.4	2.26
4-amino-2-nitrophenol	119-34-6	-0.036	0.056	3	0	0.4	
4-aminophenol	123-30-8	-0.013	0.053	2	0	0.6	2.07
4-aminopyridine	504-24-5	-0.406	0.086	1	0	0.6	1.44
4-chlorocatechol	2138-22-9	0.380	0.038	3	0	0.6	
4-ethyl-1,1'-biphenyl	5707-44-8	0.172	0.043	3	0	0.5	2.48
4-hydroxybenzoic acid	99-96-7	-1.611	0.113	2	1	0.4	0.17
4-methylbenzoic acid	99-94-5	-1.195	0.080	2	1	0.6	0.33
4-vinylpyridine	100-43-6	-0.614	0.095	1	0	0.5	2.02
5-ethylidene-8,9,10-trinorborn-2-ene	16219-75-3	-0.591	0.134	0	0	0.7	1.23
6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	91-53-2	1.212	0.101	3	0	1	1.51
6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	2439-01-2	-0.036	0.056	3	0	0.4	
8-hydroxyquinoline	148-24-3	-0.406	0.086	1	0	0.6	1.76
Acetophenone	98-86-2	-0.614	0.095	1	0	0.5	
Acetylsalicylic acid	50-78-2	-1.195	0.080	2	1	0.6	
alpha,alpha-dichlorotoluene	98-87-3	-0.614	0.095	1	0	0.5	0.85
Anisol	100-66-3	-0.614	0.095	1	0	0.5	
Atrazine-deisopropyl	1007-28-9	0.796	0.053	3	0	0.8	
Benazolin ethyl	25059-80-7	0.796	0.053	3	0	0.8	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Benzaldehyde	100-52-7	-0.614	0.095	1	0	0.5	
Benzenethiol	108-98-5	-0.614	0.095	1	0	0.5	4.09
Benzoyl-chloride	98-88-4	-0.614	0.095	1	0	0.5	
Benzyl-chloroformiate	501-53-1	-0.406	0.086	1	0	0.6	
Betariboacetate	13035-61-5	0.034	0.179	0	0	1	
Biphenyl	92-52-4	-0.221	0.060	2	0	0.5	1.60
Bis(2-chloroethyl) ether	111-44-4	-1.007	0.146	0	0	0.5	
Bismerthiazol	79319-85-0	-0.799	0.136	0	0	0.6	
Buparvaquone	88426-33-9	0.819	0.110	2	0	1	
But-3-en-3-olide	674-82-8	0.034	0.179	0	0	1	0.98
Butachlor	23184-66-9	1.212	0.101	3	0	1	3.05
Butylbenzyl phthalate	85-68-7	0.796	0.053	3	0	0.8	
Captopril	62571-86-2	-1.148	0.174	0	1	1	
Chlorohydroquinone	615-67-8	0.380	0.038	3	0	0.6	
Cinmethylin	87818-31-3	0.819	0.110	2	0	1	
Cotinine	486-56-6	0.010	0.094	1	0	0.8	
Cyclohexanone oxime	100-64-1	0.034	0.179	0	0	1	
Cyclohexanone	108-94-1	-0.175	0.156	0	0	0.9	
Cyclosulfamuron	136849-15-5	1.789	0.102	5	0	0.9	
Diallyl phthalate	131-17-9	-0.013	0.053	2	0	0.6	2.75
Diethyl disulfide	110-81-6	-1.007	0.146	0	0	0.5	1.01
Diethyl malonate	105-53-3	0.034	0.179	0	0	1	0.61
Diethyl phthalate	84-66-2	0.819	0.110	2	0	1	
Diisobutyl phthalate	84-69-5	0.819	0.110	2	0	1	1.92
Dimethyl disulphide	624-92-0	-0.175	0.156	0	0	0.9	1.93
Dimethylformamide	68-12-2	-0.175	0.156	0	0	0.9	
Dimethylnitrosamine	62-75-9	-0.175	0.156	0	0	0.9	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Dinitramine	29091-05-2	1.188	0.057	4	0	0.8	
Diphenylpropanediol	0.403	0.065	2	0	0.8		
Droperidol	548-73-2	1.188	0.057	4	0	0.8	
Enrofloxacin	93106-60-6	0.423	0.125	4	1	1	
Ethanethiol	75-08-1	-0.799	0.136	0	0	0.6	1.45
Flumazenil	78755-81-4	1.581	0.079	5	0	0.8	
Glycidyl methacrylate	106-91-2	0.034	0.179	0	0	1	1.71
Haloxyfop R	72619-32-0	1.581	0.079	5	0	0.8	
Hexamethylene diacrylate	13048-33-4	0.034	0.179	0	0	1	**2.77**
Hydrogenatedbisphenol A	80-04-6	-0.799	0.136	0	0	0.6	1.10
Hydroquinone	123-31-9	-0.013	0.053	2	0	0.6	
Isoprene	78-79-5	-1.215	0.164	0	0	0.4	0.66
Maleic anhydride	108-31-6	-1.215	0.164	0	0	0.4	
Medazepam	2898-12-6	1.188	0.057	4	0	0.8	
Methacrylonitrile	126-98-7	-0.799	0.136	0	0	0.6	
Methyl isothiocyanate	556-61-6	-0.799	0.136	0	0	0.6	2.78
Methylhydrazine	60-34-4	-0.799	0.136	0	0	0.6	2.08
m-toluic acid	99-04-7	-1.195	0.080	2	1	0.6	0.22
n-(*tert*-butyl)-2-benzothiazolylsulfenamide	95-31-8	0.010	0.094	1	0	0.8	2.23
n,*n*-dimethylhydrazine	57-14-7	-0.591	0.134	0	0	0.7	
n-cyclohexyl-2-benzothiazolylsulfenamide	95-33-0	0.010	0.094	1	0	0.8	2.10
Nitroglycerin	55-63-0	-0.591	0.134	0	0	0.7	
n-methyl- *n*,*n*-bis(2-dimethylaminoethyl)amine	3030-47-5	-0.799	0.136	0	0	0.6	
o-acetoacetotoluidide	93-68-5	0.403	0.065	2	0	0.8	
o-chlorobenzonitrile	873-32-5	-0.637	0.099	2	0	0.3	0.57
Octanedinitrile	629-40-3	-0.383	0.141	0	0	0.8	-0.61
Olaquindox	23696-28-8	0.403	0.065	2	0	0.8	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Ondansetron	99614-02-5	0.796	0.053	3	0	0.8	
o-tolidine	119-93-7	1.558	0.092	6	0	0.6	1.21
Pentane-1-thiol	110-66-7	-1.423	0.191	0	0	0.3	**1.90**
Perfluorooctane sulfonic acid	1763-23-1	-1.772	0.145	0	1	0.7	
Phenol,4,4',4''-ethylidynetris-	27955-94-8	1.558	0.092	6	0	0.6	
Phthalic anhydride	85-44-9	-0.429	0.075	2	0	0.4	
Phthalonitrile	91-15-6	-0.429	0.075	2	0	0.4	0.75
Pivaloyl chloride	3282-30-2	-0.175	0.156	0	0	0.9	
p-methoxybenzaldehyde	123-11-5	-0.013	0.053	2	0	0.6	0.53
p-phenylphenol	92-69-3	0.172	0.043	3	0	0.5	1.70
Propyl gallate	121-79-9	1.604	0.109	4	0	1	
Propyzamide	23950-58-5	0.796	0.053	3	0	0.8	
Pyrazosulfuron ethyl	93697-74-6	1.789	0.102	5	0	0.9	
Quinoline	91-22-5	0.195	0.055	2	0	0.7	0.30
Secobarbital	76-73-3	0.034	0.179	0	0	1	
Simazine	122-34-9	1.212	0.101	3	0	1	
Sulfaquinoxaline	59-40-5	-0.036	0.056	3	0	0.4	
tert-butyl 2-ethylperoxyhexanoate	3006-82-4	-0.383	0.141	0	0	0.8	1.67
tert-butylhydroperoxide	75-91-2	-0.799	0.136	0	0	0.6	-0.02
Tetrachlorocatechol	1198-55-6	1.142	0.100	6	0	0.4	
Tetrachlorohydroquinone	87-87-6	1.142	0.100	6	0	0.4	
Tetrachlorophthalic anhydride	117-08-8	1.142	0.100	6	0	0.4	
Tetrahydromethylphthalic anhydride	11070-44-3	-0.383	0.141	0	0	0.8	
Thiopental	76-75-5	0.034	0.179	0	0	1	
Thiophene	110-02-1	-1.007	0.146	0	0	0.5	0.43
Thiosemicarbazide	79-19-6	-0.591	0.134	0	0	0.7	0.64
Thiourea dioxide	4189-44-0	-1.007	0.146	0	0	0.5	

Table B1. (Continued).

NAME	CAS	Pred. $pEC_{50,NR[PLHC-1]}$ by Eq.4.3 (mM)	Hat value (*h**=0.308)	NaasC	CATS2D_01_DN	DLS_04	pLC_{50} (mM)*
Thiourea	62-56-6	-0.799	0.136	0	0	0.6	
Triclosan	3380-34-5	1.142	0.100	6	0	0.4	2.64
Trifluralin	1582-09-8	1.604	0.109	4	0	1	
Trimethylquinone	935-92-2	-0.383	0.141	0	0	0.8	
Triphenyl phosphate	115-86-6	0.380	0.038	3	0	0.6	2.40
Tris-(2,3-dibromopropyl) phosphate	126-72-7	-0.175	0.156	0	0	0.9	2.56
Tris(2-chloroethyl) phosphate	115-96-8	0.034	0.179	0	0	1	

*Bold values indicate pLC_{50} data with very weak relationship with the $pEC_{50,NR[PLHC-1]}$ prediction.

APPENDIX C: EXTERNAL SET CHEMICALS FOR $pEC_{50,NR[GFS]}$ MODEL

Table C1. External set chemicals for $pEC_{50,NR[GFS]}$ model.

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h^*=0.111)	GATS1v	DP07	pLC_{50} (mM)
5-fluorocytosine	2022-85-7	0.187	0.024	0.730	2.650	
5-fluorouracil	51-21-8	0.124	0.023	0.756	2.625	
Acetaminophen	103-90-2	0.752	0.016	0.921	5.181	
Acridine	260-94-6	1.220	0.020	0.762	5.556	
Acrivastine	87848-99-5	2.069	0.072	0.985	9.013	
Acyclovir	59277-89-3	0.737	0.018	0.980	5.463	
Alprenolol	13655-52-2	0.669	0.035	1.188	6.411	
Amitriptyline	50-48-6	1.039	0.028	1.015	6.450	
Amobarbital	57-43-2	0.264	0.027	1.213	5.477	
Ampicillin	69-53-4	1.875	0.049	0.893	8.000	
Antipyrine	60-80-0	0.641	0.017	1.001	5.322	
Bumetanide	28395-03-1	1.752	0.050	0.951	7.989	
Bupropion	34841-39-9	0.730	0.024	1.068	5.921	
Caffeine	58-08-2	0.041	0.017	1.138	4.481	
Capecitabine	154361-50-9	1.956	0.098	1.153	9.625	
Cefurexime axetil	64544-07-6	2.280	0.086	0.983	9.560	
Cephalexin	15686-71-2	2.308	0.065	0.840	8.858	
Chloramphenicol	56-75-7	1.683	0.033	0.813	7.057	
Chlorpromazine	50-53-3	1.592	0.039	0.919	7.392	
Chlorpropamide	94-20-2	1.288	0.032	0.969	6.860	
Cimetidine	51481-61-9	1.109	0.035	1.055	6.853	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h***h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Ciprofloxacin	85721-33-1	1.942	0.058	0.943	8.447	
Clonazepam	1622-61-3	1.712	0.031	0.739	6.733	
Clonidine	4205-90-7	1.174	0.019	0.752	5.380	
Cytarabine	147-94-4	1.026	0.023	0.963	6.134	
Deoxytetracycline	564-25-0	2.189	0.060	0.856	8.628	
Desipramine	50-47-5	0.859	0.024	1.025	6.030	
Doxycycline	564-25-0	2.207	0.061	0.856	8.677	
Estradiol	50-28-2	1.471	0.045	1.019	7.615	
Ethinylestradiol	57-63-6	2.015	0.052	0.866	8.224	
Etoposide	33419-42-0	1.990	0.068	0.989	8.826	
Famciclovir	104227-87-4	1.750	0.078	1.126	8.934	
Famotidine	76824-35-6	1.374	0.025	0.836	6.365	
Finasteride	98319-26-7	1.542	0.083	1.213	8.856	
Fleroxacin	79660-72-3	1.935	0.071	1.026	8.879	
Florfenicol	73231-34-2	1.673	0.040	0.901	7.510	
Flunitrazepam	1622-62-4	1.618	0.030	0.786	6.739	
Flutamide	13311-84-7	1.409	0.041	1.007	7.387	
Furaltadone	139-91-3	1.605	0.066	1.109	8.458	
Fusidic acid	6990-06-3	1.568	0.077	1.178	8.734	
Gemcitabine	95058-81-4	0.981	0.022	0.954	5.968	
Glibenclamid	10238-21-8	2.039	0.071	0.987	8.944	
Hydrochlorothiazide	58-93-5	1.402	0.024	0.771	6.087	
Hydrocortisone	50-23-7	1.792	0.071	1.074	8.762	
Imipramine	50-49-7	0.913	0.030	1.078	6.459	
Indomethacin	53-86-1	2.279	0.064	0.848	8.823	
Isoniazid	54-85-3	0.556	0.014	0.818	4.104	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h^*=0.111)	GATS1v	DP07	pLC_{50} (mM)
Isotretinoin	4759-48-2	1.937	0.097	1.155	9.586	
Ketoprofen	22071-15-4	1.725	0.039	0.869	7.472	
Lamotrigine	84057-84-1	1.910	0.034	0.558	6.274	
Levofloxacin	100986-85-4	1.963	0.071	1.016	8.899	
Lidocaine	137-58-6	0.405	0.033	1.233	5.959	
Lomefloxacin	98079-51-7	1.901	0.065	1.004	8.672	
L-tryptophan	73-22-3	1.040	0.018	0.824	5.417	
Metaxalone	1665-48-1	0.470	0.018	1.065	5.218	
Metformin	657-24-9	-0.574	0.023	1.285	3.651	
Methotrexate	59-05-2	2.335	0.074	0.895	9.226	
Methylprednisolone	83-43-2	1.589	0.055	1.048	8.084	
Metipranolol	22664-55-7	1.072	0.060	1.235	7.734	
Metoprolol	37350-58-6	1.277	0.077	1.261	8.416	
Metronidazole	443-48-1	-0.089	0.016	1.134	4.115	
Midazolam	59467-70-8	1.688	0.033	0.800	7.001	
Nadolol	42200-33-9	1.238	0.053	1.147	7.695	
Naproxen	22204-53-1	1.596	0.040	0.931	7.468	
Nicotine	22083-74-5	-0.113	0.015	1.131	4.035	
Nisoldipine	63675-72-9	1.257	0.042	1.068	7.315	
Nitrofurazone	59-87-0	1.300	0.022	0.802	5.984	
Norfloxacin	70458-96-7	1.828	0.059	0.990	8.403	
Ofloxacin	82419-36-1	1.963	0.071	1.016	8.899	
Oxprenolol	6452-71-7	0.989	0.050	1.199	7.318	
Oxyphenbutazone	129-20-4	1.533	0.041	0.962	7.469	
Oxytetracycline	79-57-2	2.307	0.064	0.837	8.837	
Paracetamol	103-90-2	0.752	0.016	0.921	5.181	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Paroxetine	61869-08-7	1.771	0.053	0.969	8.136	
Pentobarbital	76-74-4	0.404	0.030	1.213	5.847	
Phenobarbital	50-06-6	0.806	0.016	0.904	5.233	
Phenoxymethylpenicillinic Acid	87-08-1	1.617	0.041	0.934	7.539	
Phenylbutazone	50-33-9	1.369	0.037	0.984	7.155	
Phenytoin	57-41-0	1.246	0.021	0.790	5.776	
Pindolol	13523-86-9	1.304	0.046	1.078	7.495	
Piperazine	110-85-0	-2.019	0.049	1.403	0.471	
Prednisolone	50-24-8	1.800	0.063	1.028	8.533	
Procaine	59-46-1	1.077	0.052	1.188	7.490	
Progesterone	57-83-0	1.513	0.065	1.132	8.340	
Promazine	58-40-2	1.260	0.032	0.977	6.829	
Propylthiouracil	51-52-5	0.266	0.014	1.039	4.537	
Quinidine	56-54-2	1.439	0.043	1.017	7.521	
Quinine	130-95-0	1.231	0.035	1.017	6.971	
Ranitidine	66357-35-5	0.937	0.042	1.166	7.001	
Sotalol	3930-20-9	1.239	0.053	1.145	7.687	
Sulfaguanidine	57-67-0	1.303	0.022	0.808	6.026	
Sulfaphenazole	526-08-9	1.728	0.036	0.830	7.268	
Sumatriptan	103628-46-2	1.499	0.060	1.111	8.190	
Terbinafine	91161-71-6	1.331	0.044	1.059	7.463	
Testosterone	58-22-0	1.245	0.049	1.120	7.567	
Theophylline	58-55-9	0.083	0.013	1.033	4.021	
Timolol	26839-75-8	1.141	0.064	1.237	7.925	
Tolazamide	1156-19-0	1.499	0.060	1.106	8.162	
Tolbutamide	64-77-7	0.987	0.038	1.117	6.868	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Trenbolone acetate	10161-34-9	1.919	0.071	1.031	8.864	
Triamterene	396-01-0	2.109	0.042	0.688	7.506	
Triflupromazine	146-54-3	1.670	0.048	0.968	7.865	
Warfarin	81-81-2	1.762	0.040	0.859	7.515	
2,6-dichlorobenzonitrile	1194-65-6	0.678	0.028	0.602	3.254	
Acetochlor	34256-82-1	0.422	0.024	1.145	5.525	
Acroleine	107-02-8	-1.532	0.038	1.190	0.604	
Ametryn	834-12-8	0.757	0.037	1.178	6.590	
Anilofos	64249-01-0	1.756	0.055	0.992	8.223	
Azocyclotin	41083-11-8	0.792	0.045	1.221	6.915	
Benalaxyl	71626-11-4	1.286	0.043	1.062	7.360	
Benomyl	17804-35-2	1.178	0.040	1.079	7.165	
Bensulfuron-methyl	83055-99-6	2.250	0.095	1.040	9.789	
Bispyribac	125401-75-4	2.069	0.078	1.020	9.201	
Bromoxynil	1689-84-5	1.452	0.036	0.463	4.547	
Chloridazon	1698-60-8	1.533	0.026	0.721	6.162	
Chlorimuron-ethyl	90982-32-4	2.229	0.076	0.945	9.219	
Chlorothalonil	1897-45-6	1.686	0.036	0.468	5.192	
Chlorotoluron	15545-48-9	1.064	0.029	1.021	6.551	
Chlorsulfuron	64902-72-3	1.738	0.045	0.922	7.795	
Cyanazine	21725-46-2	0.786	0.023	1.044	5.939	
Cyhexatin	13121-70-5	0.469	0.039	1.261	6.279	
Decanoic acid	334-48-5	-0.699	0.037	1.444	4.184	
Dimethachlon	24096-53-5	1.282	0.021	0.754	5.677	
Endosulfan	115-29-7	1.322	0.022	0.667	5.311	
Ethalfluraline	55283-68-6	0.876	0.029	1.077	6.356	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Ethametsulfuron	111353-84-5	1.988	0.066	0.975	8.744	
Ethoxyquin	91-53-2	0.831	0.030	1.099	6.356	1.51
Fenamiphos	22224-92-6	1.085	0.050	1.172	7.424	
Fenhexamid	126833-17-8	1.531	0.035	0.902	7.138	
Fenpropidin	67306-00-7	1.506	0.093	1.269	9.064	
Fentin hydroxide	76-87-9	1.436	0.037	0.960	7.202	
Flumetsulam	98967-40-9	2.142	0.062	0.889	8.684	
Fluroxypyr	69377-81-7	1.678	0.029	0.616	5.975	
Glyphosate	1071-83-6	-0.367	0.015	1.102	3.206	
Iprodione	36734-19-7	2.022	0.056	0.894	8.395	
Cybutryne	28159-98-0	1.050	0.041	1.122	7.061	
Isoproturon	34123-59-6	1.025	0.051	1.193	7.380	
Lindane	58-89-9	0.729	0.024	0.632	3.552	
Mefenacet	73250-68-7	2.210	0.063	0.868	8.750	
Metalaxyl	57837-19-1	0.509	0.033	1.212	6.119	
Methabenzthiazuron	18691-97-9	1.384	0.030	0.903	6.755	
Metolachlor	51218-45-2	0.546	0.029	1.165	5.963	
Metribuzin	21087-64-9	0.547	0.020	1.072	5.459	
Metsulfuron-methyl	74223-64-6	1.698	0.056	1.016	8.199	
Molinate	2212-67-1	0.340	0.033	1.248	5.867	
Nicosulfuron	111991-09-4	1.666	0.064	1.075	8.435	
Oxadiargyl	39807-15-3	1.900	0.050	0.897	8.087	
Oxadiazon	19666-30-9	1.538	0.045	0.996	7.667	
Oxadixyl	77732-09-3	0.465	0.024	1.136	5.590	
Pendimethalin	40487-42-1	0.705	0.031	1.146	6.279	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Pretilachlor	51218-49-6	0.674	0.037	1.199	6.486	2.11
Prometryn	7287-19-6	0.884	0.046	1.201	7.050	
Quinclorac	84087-01-4	1.596	0.028	0.598	5.660	
Tebuthiuron	34014-18-1	0.769	0.036	1.168	6.569	
Terbumeton	33693-04-8	0.456	0.049	1.328	6.608	
Terbuthylazine	5915-41-3	0.542	0.028	1.163	5.941	
Terbutryn	886-50-0	0.751	0.040	1.201	6.699	
Thiobencarb	28249-77-6	1.348	0.038	1.007	7.226	2.30
Thiophanate methyl	23564-05-8	1.796	0.042	0.861	7.617	
Tribenuron	106040-48-6	1.600	0.050	1.009	7.903	
1,1,2,2-tetrachloroethane	79-34-5	0.137	0.044	0.580	1.705	
1,1,2-trichloroethane	79-00-5	-0.228	0.039	0.698	1.381	
1,1-dichloroethane	75-34-3	-1.020	0.045	0.888	0.318	
1,1-dichloroethylene	75-35-4	-0.575	0.059	0.669	0.306	0.33
1,2,3-trichloropropane	96-18-4	-0.223	0.025	0.815	2.029	
1,2-dibromo-3-chloropropane	96-12-8	-0.056	0.033	0.700	1.846	0.78
1,2-dichloroethane	107-06-2	-0.627	0.030	0.888	1.357	
1,2-dichloropropane	78-87-5	-0.867	0.028	0.998	1.320	-0.15
1,3-dibromopropane	109-64-8	-0.521	0.033	0.823	1.285	1.41
1,3-dichloropropane	142-28-9	-0.836	0.027	0.998	1.401	
1,3-dichloropropene	542-75-6	-0.090	0.022	0.817	2.392	1.87
1-chlorobutane	109-69-3	-1.101	0.027	1.331	2.509	
3,4-dichlorobut-1-ene	760-23-6	-0.588	0.026	0.921	1.639	0.67
Acetic acid, bromo-, 2-butene-1,4-diyl ester	20679-58-7	0.699	0.015	0.922	5.046	
Bromodichloromethane	75-27-4	-0.172	0.071	0.511	0.514	0.77
Cyclohexane	110-82-7	-2.032	0.049	1.417	0.514	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Dibromochloromethane	124-48-1	-0.114	0.070	0.501	0.611	0.42
Ethyl bromide	74-96-4	-1.353	0.043	1.031	0.213	
Ethylcyclohexane	1678-91-7	-1.167	0.034	1.438	2.915	2.18
Methylcyclohexane	108-87-2	-1.879	0.045	1.429	0.982	1.67
Pentachloroethane	76-01-7	0.289	0.051	0.509	1.721	
Tetrachloroethene	127-18-4	-1.762	0.045	1.500	1.677	1.07
Trichloroethylene	79-01-6	0.081	0.055	0.534	1.306	0.54
1-(*n*-phenylamino)-naphthalene	90-30-2	1.700	0.033	0.799	7.026	2.50
1,2-dimethylnaphthalene	573-98-8	0.385	0.013	0.920	4.207	1.83
1,3-dimethylnapthalene	575-41-7	0.394	0.013	0.920	4.230	1.94
1,5-napthalenediamine	2243-62-1	0.638	0.018	0.724	3.812	0.97
1,8-naphthylenediamine	479-27-6	0.537	0.019	0.724	3.543	1.45
1-methylnaphthalene	90-12-0	0.162	0.015	0.868	3.335	1.40
2,7-dimethylnaphthalene	582-16-1	0.669	0.015	0.920	4.957	1.94
2-methylnaphthalene	91-57-6	0.465	0.013	0.868	4.134	1.87
3-hydroxy-2-naphthoic acid	92-70-6	1.294	0.021	0.709	5.465	
Isopropylnaphthalene	29253-36-9	0.375	0.013	0.964	4.419	**2.36**
n-phenyl-2-naphthylamine	135-88-6	1.810	0.037	0.799	7.318	
ß-Naphthol	135-19-3	0.642	0.016	0.761	4.023	1.56
ß-Naphthylamine	91-59-8	0.656	0.016	0.762	4.066	1.56
Tetralin	119-64-2	-0.206	0.014	1.004	3.099	1.34
1,2,3-trichlorobenzene	87-61-6	0.593	0.034	0.565	2.828	1.75
1,2-dichlorobenzene	95-50-1	0.171	0.034	0.647	2.158	1.59
1,3-dichlorobenzene	541-73-1	0.376	0.029	0.647	2.701	1.41
1,4-benzenediamine, *n*,*n*'-bis(1-methylpropyl)-	101-96-2	1.038	0.064	1.262	7.790	2.78
1,4-dichlorobenzene	106-46-7	0.582	0.025	0.647	3.246	1.82

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07	pLC_{50} (mM)
2,4,6-trichlorophenylhydrazine	5329-12-4	1.342	0.036	0.468	4.284	2.33
2,4,6-trinitrotoluene	118-96-7	0.980	0.018	0.867	5.493	
2,4-diamino-6-nitrotoluene	6629-29-4	0.501	0.014	0.826	4.003	
2,4-diaminotoluene	95-80-7	-0.034	0.015	0.896	2.968	
2,4-dichlorotoluene	95-73-8	0.435	0.019	0.736	3.339	
2,4-dinitrotoluene	121-14-2	0.681	0.014	0.902	4.892	
2,5-diaminotoluene	95-70-5	-0.039	0.016	0.896	2.954	**2.81**
2,5-dichlorotoluene	19398-61-9	0.458	0.019	0.736	3.400	1.60
2,6-diamino-4-nitrotoluene	59229-75-3	0.673	0.014	0.826	4.458	
2,6-diaminotoluene	823-40-5	-0.253	0.019	0.896	2.389	
2,6-dichlorotoluene	118-69-4	0.458	0.019	0.736	3.400	1.85
2,6-dinitrotoluene	606-20-2	0.610	0.014	0.902	4.702	0.73
2-amino-4,6-dinitrotoluene	35572-78-2	0.876	0.016	0.827	4.999	
2-amino-4-nitrotoluene	99-55-8	0.503	0.013	0.880	4.301	
2-amino-6-nitrotoluene	603-83-8	0.224	0.014	0.880	3.562	
2-chlorohydroquinonedimethylether	2100-42-7	0.454	0.014	0.980	4.713	0.79
2-chlorotoluene	95-49-8	-0.291	0.024	0.848	2.028	1.22
2-methyl-4-chlorophenoxyacetic acid	94-74-6	0.991	0.018	0.878	5.580	
2-phenylpropene	98-83-9	-0.235	0.014	1.000	3.001	1.21
3,4-dichlorotoluene	95-75-0	0.419	0.019	0.736	3.298	1.57
3,5-bis(trifluoromethyl)benzylamine	85068-29-7	0.970	0.018	0.893	5.607	1.34
3-chlorotoluene	108-41-8	-0.101	0.019	0.848	2.530	
4,6-dichlororesorcinol	137-19-9	0.971	0.036	0.499	3.471	
4-allyl-1,2-dimethoxybenzene	93-15-2	0.524	0.025	1.131	5.719	1.10
4-chlororesorcinol	95-88-5	0.558	0.028	0.622	3.047	
4-chlorotoluene	106-43-4	0.099	0.016	0.848	3.060	1.32

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
4-nitrotoluene-2-sulphonic acid	121-03-9	0.787	0.016	0.914	5.237	
4-*tert*-Butyltoluene	98-51-1	-0.042	0.019	1.182	4.500	1.94
4-toluenesulfonyl chloride	98-59-9	0.675	0.014	0.860	4.648	0.39
Atenolol	29122-68-7	1.134	0.047	1.141	7.387	
Benzalacetone	122-57-6	0.825	0.019	0.964	5.608	1.41
Benzotrifluoride	98-08-8	0.649	0.024	0.651	3.444	1.01
Benzyl alcohol	100-51-6	-0.288	0.017	0.940	2.535	
Benzyl cyanide	140-29-4	0.187	0.014	0.888	3.509	
Benzylamine	100-46-9	-0.289	0.017	0.945	2.560	
Bromobenzene	108-86-1	-0.019	0.032	0.698	1.933	1.56
Butylbenzene	104-51-8	0.315	0.022	1.150	5.269	1.61
Catechol	120-80-9	-0.263	0.030	0.776	1.712	
Chlorobenzene	108-90-7	-0.232	0.031	0.758	1.696	1.23
Cyclohexylbenzene	827-52-1	0.392	0.018	1.086	5.126	2.13
Diisopropylbenzene	25321-09-9	-0.372	0.018	1.208	3.769	**2.30**
Dimethyl phthalate	131-11-3	0.388	0.015	1.036	4.845	
Diuron	330-54-1	1.432	0.029	0.880	6.758	
Divinylbenzene	1321-74-0	-0.030	0.014	0.950	3.273	1.49
Ethylbenzene	100-41-4	-0.538	0.017	1.063	2.542	
Fenobucarb	3766-81-2	0.174	0.020	1.157	4.934	1.33
Isopropylbenzene	98-82-8	-0.468	0.016	1.111	2.988	
Metaxylene hexafluoride	402-31-3	0.749	0.016	0.923	5.186	1.47
m-phenylenebis(methylamine)	1477-55-0	0.256	0.012	0.969	4.131	0.19
m-phenylenediamine	108-45-2	-0.129	0.024	0.799	2.189	
n-propylbenzene	103-65-1	-0.029	0.015	1.111	4.149	
o-phenylenediamine	95-54-5	-0.276	0.028	0.799	1.801	1.37

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
o-toluenesulfonamide	88-19-7	0.070	0.013	0.938	3.470	
Oxyfluorfen	42874-03-3	2.240	0.066	0.881	8.899	
p-cymene	99-87-6	-0.074	0.017	1.150	4.242	**1.83**
p-phenylenediamine	106-50-3	0.068	0.020	0.799	2.710	3.21
Propoxur	114-26-1	0.206	0.020	1.147	4.966	
Resorcinol	108-46-3	-0.138	0.027	0.776	2.041	
Sulphanilamide	63-74-1	0.728	0.015	0.815	4.544	
Thiamphenicol	15318-45-3	2.121	0.060	0.886	8.613	
Toluene	108-88-3	-0.818	0.026	1.000	1.460	0.57
Trans-cinnamic acid	140-10-3	1.036	0.018	0.856	5.581	
2-(1,1-dimethyl)-4,6-dimethylphenol	1879-09-0	-0.056	0.018	1.171	4.402	**1.85**
2,3,4,5-tetrachlorophenol	4901-51-3	1.391	0.046	0.378	3.923	
2,3,5,6-tetrachlorophenol	935-95-5	1.422	0.046	0.378	4.005	
2,3,5-trichlorophenol	933-78-8	1.079	0.037	0.476	3.631	
2,3,5-trimethylphenol	697-82-5	-0.255	0.014	1.063	3.289	
2,3,6-trichlorophenol	933-75-5	1.042	0.038	0.476	3.534	
2,3,6-trimethylphenol	2416-94-6	-0.288	0.014	1.063	3.202	
2,3-dimethylphenol	526-75-0	-0.477	0.018	1.009	2.409	
2,4,6-tribromophenol	118-79-6	1.287	0.041	0.426	3.909	2.34
2,4,6-trimethylphenol	527-60-6	-0.262	0.014	1.063	3.271	
2,4-dibromophenol	615-58-7	0.957	0.034	0.512	3.504	1.84
2,4-di-tert-butylphenol	96-76-4	0.248	0.027	1.217	5.457	2.48
2,5-dimethylphenol	95-87-4	-0.245	0.014	1.009	3.024	1.33
2,6-dimethylphenol	576-26-1	-0.465	0.017	1.009	2.442	0.91
2,6-di-sec-butylphenol	5510-99-6	0.615	0.038	1.217	6.427	3.14
2-allylphenol	1745-81-9	0.186	0.012	0.951	3.849	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
2-ethylphenol	90-00-6	-0.393	0.016	1.009	2.633	
2-methoxyphenol	90-05-1	-0.324	0.016	0.977	2.642	
2-*n*-propylphenol	644-35-9	0.063	0.014	1.063	4.132	
2-*tert*-butyl phenol	88-18-6	-0.330	0.015	1.106	3.325	**1.61**
2-*tert*-butyl-*p*-cresol	2409-55-4	-0.124	0.016	1.141	4.061	**1.94**
3,4,5-trichloroguaiacol	57057-83-7	1.008	0.023	0.632	4.289	
3,4,5-trichlorophenol	609-19-8	1.028	0.038	0.476	3.495	
3,4-dimethylphenol	95-65-8	-0.284	0.015	1.009	2.921	
3-ethylphenol	620-17-7	-0.204	0.014	1.009	3.131	
3-trifluoromethyl-4-nitrophenol	88-30-2	0.703	0.015	0.792	4.353	
4-(1-methylethenyl)phenol	4286-23-1	0.296	0.012	0.951	4.138	1.16
4-(2,4-dichlorophenoxy)-phenol	40843-73-0	1.978	0.037	0.682	7.127	
4,5,6-trichloroguaiacol	2668-24-8	1.075	0.023	0.632	4.468	
4,5-dichloroguaiacol	2460-49-3	0.806	0.018	0.721	4.239	
4-chloro-2-methylphenol	1570-64-5	0.248	0.018	0.793	3.154	
4-chloro-2-nitrophenol	89-64-5	0.829	0.023	0.643	3.877	1.20
4-ethylphenol	123-07-9	0.006	0.013	1.009	3.687	
4-*n*-octylphenol	1806-26-4	1.758	0.098	1.217	9.450	3.37
4-pentylphenol	14938-35-3	1.016	0.042	1.141	7.074	2.07
6-*tert*-butyl-*m*-cresol	88-60-8	0.042	0.017	1.141	4.498	1.78
6-*tert*-butyl-*o*-cresol	2219-82-1	-0.159	0.016	1.141	3.967	1.58
Butylated hydroxyanisole	25013-16-5	0.186	0.020	1.158	4.972	1.49
Methyl *p*-hydroxybenzoate	99-76-3	0.376	0.013	0.937	4.275	0.40
o-*sec*-butylphenol	89-72-5	-0.012	0.015	1.106	4.167	1.40
p-*sec*-butylphenol	99-71-8	0.076	0.015	1.106	4.399	1.76
p-*tert*-butylphenol	98-54-4	0.066	0.015	1.106	4.371	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h^*=0.111)	GATS1v	DP07	pLC_{50} (mM)
Tetrachloroguaiacol	2539-17-5	1.263	0.028	0.557	4.558	
Thymol	89-83-8	0.018	0.015	1.106	4.246	1.50
Trichlorosyringol	2539-26-6	0.997	0.018	0.742	4.858	
1-nitropyrene	5522-43-0	1.692	0.029	0.689	6.409	
2-acetamidophenoxazin-3-one	1916-55-8	1.982	0.045	0.809	7.826	
2-amino-7-methoxyphenoxazin-3-one		1.949	0.042	0.791	7.642	
2-aminophenoxazin-3-one	1916-59-2	1.713	0.030	0.706	6.556	
9-vinylcarbazole	1484-13-5	0.941	0.016	0.819	5.127	
Acenaphthene	83-32-9	0.296	0.016	0.802	3.330	
Dibenzo[b,f]cyclohepten-1-one	2222-33-5	1.188	0.019	0.755	5.433	
Dibenzothiophene	132-65-0	1.171	0.021	0.679	4.975	2.12
Fluorene	86-73-7	0.942	0.017	0.778	4.907	
Phenanthrene	85-01-8	1.032	0.018	0.758	5.038	2.10
Phenothiazine	92-84-2	1.440	0.024	0.692	5.759	2.41
1-decanol	112-30-1	-0.321	0.062	1.546	5.738	1.75
1-nonanol	143-08-8	0.481	0.099	1.551	7.885	1.65
2-(2-butoxyethoxy)ethanol	112-34-5	-0.797	0.053	1.566	4.587	
2-butoxyethanol	111-76-2	-1.564	0.049	1.579	2.630	
2-isoproxyethanol	109-59-1	-1.920	0.054	1.593	1.765	
3-pentanol	71-41-0	-2.111	0.057	1.590	1.243	
Cyclohexanol	108-93-0	-1.512	0.034	1.340	1.470	
Hexanol	111-27-3	-1.749	0.050	1.575	2.118	
Isodecyl alcohol	25339-17-7	-0.892	0.048	1.546	4.228	1.43
1,6-hexanediamine	124-09-4	-0.095	0.061	1.506	6.118	0.21
2-(dibutylamino)ethanol	102-81-8	-0.189	0.076	1.589	6.321	0.78

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h*=0.111)	GATS1v	DP07	pLC_{50} (mM)
2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)	6864-37-5	0.915	0.068	1.314	7.746	1.04
2-amino-2-ethylpropanediol	115-70-8	-1.464	0.035	1.405	1.949	
Cyclohexylamine	108-91-8	-1.500	0.034	1.350	1.557	0.48
Dibutylamine	111-92-2	-0.809	0.062	1.619	4.843	
Diethanolamine	111-42-2	-1.828	0.045	1.486	1.428	
Diethylamine	109-89-7	-2.575	0.078	1.728	0.765	0.43
Diethylnitrosamine	55-18-5	-2.000	0.056	1.601	1.597	
Diisopropylamine	108-18-9	-1.921	0.060	1.656	2.105	
Piperidine	110-89-4	-2.028	0.049	1.410	0.485	
t-butylamine	75-64-9	-2.546	0.070	1.621	0.260	
Triethylamine	121-44-8	-2.181	0.071	1.730	1.818	0.63
2-(dimethylamino)ethyl methacrylate	2867-47-2	-0.442	0.042	1.451	4.901	0.92
2-ethylhexyl methacrylate	688-84-6	0.332	0.062	1.422	6.792	1.85
2-hydroxyethyl acrylate	818-61-1	-0.485	0.020	1.245	3.670	1.25
2-hydroxyethyl methacrylate	868-77-9	-0.515	0.022	1.281	3.786	
Ethyl trichloroacetate	515-84-4	0.186	0.015	0.843	3.261	0.64
Ethylacrylate	140-88-5	-1.016	0.027	1.334	2.749	**1.92**
Isobutyl acetate	110-19-0	-1.266	0.040	1.503	3.004	0.83
Methyl acrylate	96-33-3	-1.282	0.029	1.303	1.876	**1.79**
Methyl methacrylate	80-62-6	-1.604	0.037	1.334	1.195	
n-butyl acrylate	141-32-2	-0.282	0.035	1.374	4.907	**1.73**
n-butyl methacrylate	97-88-1	-0.178	0.039	1.388	5.259	1.40
Testosterone propionate	57-85-2	1.850	0.094	1.170	9.438	
Tetracaine	94-24-6	1.541	0.093	1.260	9.108	
Vinyl acetate	108-05-4	-1.337	0.030	1.303	1.733	**1.55**

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
2,3,4-trichloroaniline	634-67-3	1.024	0.039	0.467	3.437	
2,3-dichloroaniline	608-27-5	0.517	0.035	0.569	2.649	
2,3-dimethylaniline	87-59-2	-0.474	0.017	1.014	2.446	
2,4,5-trichloroaniline	636-30-6	1.163	0.037	0.467	3.804	
2,4,6-trichloroaniline	634-93-5	1.097	0.038	0.467	3.629	1.57
2,4,6-trimethylaniline	88-05-1	-0.255	0.014	1.067	3.311	0.39
2,4-dichloroaniline	554-00-7	0.757	0.031	0.569	3.285	1.30
2,4-dimethylaniline	95-68-1	-0.256	0.014	1.014	3.022	
2,5-dichloroaniline	95-82-9	0.793	0.030	0.569	3.380	1.87
2,5-dimethylaniline	95-78-3	-0.245	0.014	1.014	3.051	
2,6-dichloroaniline	608-31-1	0.566	0.034	0.569	2.778	
2,6-diethylaniline	579-66-8	-0.123	0.015	1.109	3.890	
2,6-dimethylaniline	87-62-7	-0.461	0.017	1.014	2.479	
2-chloro-5-methylaniline	95-81-8	0.263	0.017	0.794	3.200	1.11
2-chloroaniline	95-51-2	0.015	0.032	0.692	1.991	1.24
2-ethylaniline	578-54-1	-0.395	0.016	1.014	2.655	
2-methyl-4-nitroaniline	99-52-5	0.483	0.013	0.880	4.247	
2-methylaniline	95-53-4	-0.562	0.023	0.945	1.839	-0.15
2-nitroaniline	88-74-4	0.178	0.019	0.784	2.921	0.52
2-nitro-*p*-anisidine	96-96-8	0.450	0.013	0.917	4.362	0.61
3,4,5-trichloroaniline	634-91-3	1.060	0.039	0.467	3.533	
3,4-dichloroaniline	95-76-1	0.729	0.031	0.569	3.209	1.17
3,4-dimethylaniline	95-64-7	-0.265	0.014	1.014	2.999	
3,5-dichloroaniline	626-43-7	0.707	0.031	0.569	3.151	
3,5-dimethylaniline	108-69-0	-0.312	0.015	1.014	2.873	0.55
3-chloroaniline	108-42-9	0.194	0.027	0.692	2.463	1.16

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
3-ethylaniline	587-02-0	-0.181	0.014	1.014	3.220	
3-methylaniline	108-44-1	-0.398	0.019	0.945	2.272	
3-nitroaniline	99-09-2	0.407	0.016	0.784	3.527	0.19
4-chloro-2-nitroaniline	89-63-4	0.847	0.023	0.637	3.891	1.01
4-chloroaniline	106-47-8	0.388	0.024	0.692	2.976	1.34
4-ethylaniline	589-16-2	0.026	0.013	1.014	3.768	
4-Fluoroaniline	371-40-4	0.003	0.021	0.804	2.567	
4-isopropylaniline	99-88-7	0.072	0.014	1.067	4.176	0.47
4-methylaniline	106-49-0	-0.205	0.016	0.945	2.782	-0.05
4-nitroaniline	100-01-6	0.592	0.015	0.784	4.016	0.21
Aniline, *p*-(phenylazo)	60-09-3	1.916	0.043	0.816	7.689	2.75
Chlorfluoroaniline	21397-08-0	0.252	0.031	0.649	2.383	
Dithiodianiline	722-27-0	1.780	0.031	0.668	6.526	
n,*n*-diethylaniline	91-66-7	-0.306	0.019	1.214	3.974	0.76
n,*n*'-dimethylaniline		-0.511	0.017	1.147	3.070	
n,*n*-dimethylaniline	121-69-7	-0.511	0.017	1.147	3.070	
n-ethylaniline	103-69-5	-0.174	0.014	1.080	3.596	0.22
p-anisidine	104-94-9	0.087	0.013	0.978	3.733	
2-ethyl butanoic acid	88-09-5	-1.357	0.034	1.410	2.259	0.14
2-methyl butanoic acid	600-07-7	-1.658	0.039	1.394	1.377	
3-methyl butanoic acid	503-74-2	-1.334	0.033	1.394	2.233	
Acrylic acid	79-10-7	-1.208	0.036	1.040	0.644	0.07
Adipic acid	124-04-9	-0.274	0.018	1.189	3.923	
A-fluoro-b-alanine	3821-81-6	-1.102	0.028	1.109	1.300	
Amidosulfonic acid	5329-14-6	-1.331	0.042	1.033	0.281	
Chloroacetic acid	79-11-8	-0.659	0.039	0.816	0.881	0.12

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Dansylglycine	1091-85-6	1.742	0.052	0.973	8.081	
Flumequine	42835-25-6	1.270	0.024	0.867	6.259	
Gentisic acid	490-79-9	0.703	0.020	0.685	3.772	
Glyoxylic acid	298-12-4	-0.960	0.042	0.893	0.503	0.26
Heptanoic acid	111-14-8	0.086	0.052	1.422	6.141	0.24
Isocyanuric acid	108-80-5	0.280	0.030	0.651	2.469	
Malonic acid diethylester	105-53-3	-0.074	0.043	1.397	5.581	0.61
Methacrylic acid	79-41-4	-1.315	0.033	1.146	0.939	
Octanoic acid	124-07-2	0.449	0.069	1.431	7.149	0.45
Orthoformic acid trimethylester	149-73-5	-2.396	0.076	1.746	1.337	
Oxolinic acid	14698-29-4	1.472	0.031	0.888	6.906	
Perfluorooctanoic acid	335-67-1	1.135	0.065	1.245	7.953	
Pivalic acid	75-98-9	-1.788	0.042	1.394	1.033	
Pyridaphenthion	119-12-0	1.315	0.038	1.013	7.171	
Sorbic acid	110-44-1	0.305	0.018	1.100	4.971	0.17
2-methylvaleraldehyde	123-15-9	-1.282	0.039	1.490	2.893	
Capronaldehyde	66-25-1	-0.857	0.041	1.491	4.022	
Crotonaldehyde	4170-30-3	-1.128	0.026	1.260	2.051	**2.99**
Glutaraldehyde	111-30-8	-0.794	0.024	1.310	3.207	**1.06**
Propionaldehyde	123-38-6	-2.105	0.052	1.485	0.689	
1,2,4-trichloro-5-nitrobenzene	89-69-0	1.360	0.030	0.522	4.623	
2-chloro-1-fluoro-4-nitrobenzene	350-30-1	0.859	0.019	0.699	4.259	**1.94**
2-chloronitrobenzene	88-73-3	0.327	0.022	0.723	2.983	
2-nitroanisole	91-23-6	-0.060	0.013	0.985	3.383	**0.51**
2-nitrotoluene	88-72-2	-0.145	0.015	0.944	2.936	
3,4-dichloronitrobenzene	99-54-7	1.121	0.024	0.615	4.497	**1.61**

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
3-chloronitrobenzene	121-73-3	0.606	0.018	0.723	3.722	
3-nitroanisole	555-03-3	0.385	0.013	0.985	4.559	0.41
4-amino-2,6-dinitrotoluene	19406-51-0	0.829	0.015	0.827	4.875	
4-amino-2-nitrotoluene	119-32-4	0.285	0.013	0.880	3.724	
4-chloro-3-methylnitrobenzene	13290-74-9	0.698	0.015	0.811	4.442	
4-chloronitrobenzene	100-00-5	0.817	0.018	0.723	4.279	
4-methylnitrobenzene	99-99-0	0.302	0.012	0.944	4.116	0.57
alpha-chloro-4-nitrotoluene	100-14-1	0.921	0.016	0.811	5.033	2.45
Nitrobenzene	98-95-3	-0.018	0.017	0.864	2.836	
1,3-diphenylguanidine	102-06-7	0.863	0.016	0.844	5.058	1.09
2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	901-44-0	2.289	0.101	1.058	9.990	**1.18**
2,3,4,4'-tetrahydroxybenzophenon	31127-54-5	2.100	0.041	0.679	7.434	0.84
2,4-diamino-6-phenyl-*s*-triazine	91-76-9	1.293	0.021	0.761	5.745	
2-hydroxy-4-methoxybenzophenone	131-57-7	1.536	0.032	0.863	6.941	1.78
4,4'-diaminodiphenyl ether	101-80-4	1.733	0.034	0.790	7.065	
4,4'-dihydroxydiphenylmethane	620-92-8	1.548	0.029	0.820	6.739	1.19
4,4'-methylenedianiline	101-77-9	1.572	0.030	0.824	6.825	0.98
Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-	51963-82-7	0.705	0.038	1.199	6.567	1.08
Benzophenone	119-61-9	1.130	0.019	0.826	5.667	
Bis(4-hydroxyphenyl)sulfone	80-09-1	1.603	0.029	0.782	6.677	
Bisphenol A	80-05-7	1.328	0.028	0.911	6.652	1.46
Dibenzyl ether	103-50-4	1.251	0.029	0.952	6.670	1.46
Dibromocresyl glycidyl ether	30171-80-3	1.060	0.018	0.778	5.221	2.39
Diphenyl ether	101-84-8	1.048	0.019	0.865	5.662	1.98
Diphenylamine	122-39-4	1.142	0.020	0.852	5.840	1.41
di-*p*-tolylamine	620-93-9	1.498	0.037	0.943	7.274	2.66

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Hydrazobenzene	122-66-7	1.304	0.024	0.850	6.255	3.63
n,*n*'-Bis(2-methylphenyl)guanidine	97-39-2	0.682	0.015	0.929	5.040	1.10
Procymidone	32809-16-8	1.560	0.029	0.799	6.656	
Pyrimethamine	58-14-0	1.597	0.030	0.796	6.737	
Styrene-7,8-oxide	96-09-3	-0.032	0.014	0.919	3.097	1.14
Sulfadiazine	68-35-9	1.805	0.041	0.847	7.566	
Sulfadimethoxine	122-11-2	1.879	0.059	0.970	8.427	
Sulfamethazine	57-68-1	1.841	0.051	0.930	8.110	
Sulfamethoxazole	723-46-6	1.865	0.046	0.872	7.858	
Tetrabromobisphenol A	79-94-7	2.310	0.046	0.603	7.576	1.77
Trimethoprim	738-70-5	1.547	0.048	1.014	7.789	
1,1'-oxybis-butane	142-96-1	-0.437	0.074	1.626	5.865	
1,2,3-trihydroxybenzene	87-66-1	0.089	0.029	0.705	2.257	
1,5-cyclooctadiene	111-78-4	-1.203	0.028	1.188	1.462	0.92
1-Benzo[b]thien-2-ylethan-1-one	22720-75-8	1.109	0.018	0.763	5.268	1.25
1-chloro-2,4-dinitrobenzene	97-00-7	1.146	0.020	0.686	4.948	3.10
1-cyclohexene-1-carbonitrile	1855-63-6	-0.418	0.015	1.074	2.919	0.42
1-mercaptooctane	111-88-6	0.264	0.066	1.456	6.797	2.65
1-methoxy-2-propanol	107-98-2	-2.023	0.057	1.614	1.605	
2-(1'-cyclohexenyl)cyclohexanone	1502-22-3	-0.020	0.019	1.174	4.515	0.79
2,2,5,5,-tetramethylhydrofuran	15045-43-9	-1.444	0.037	1.447	2.230	
2,2,6,6-tetramethylpiperidin-4-ol	2403-88-5	-1.015	0.031	1.396	3.089	
2,3,3,3,2',3',3',3'-octachlorodipropyl ether	127-90-2	1.819	0.032	0.658	6.575	2.81
2,3-dichloro-1,4-napthoquinone	117-80-6	1.267	0.024	0.607	4.838	3.86
2,4,6-trimethylbenzaldehyde	487-68-3	-0.001	0.013	1.061	3.951	1.09
2-acetoxy-1,4-naphthoquinone	1785-65-5	1.239	0.022	0.838	6.018	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
2-aminophenol	95-55-6	-0.269	0.029	0.785	1.743	**2.21**
2-aminopyridine	504-29-0	-0.591	0.032	0.859	1.295	0.93
2-butanone oxime	96-29-7	-1.844	0.044	1.431	1.086	
2-butenedinitrile, (E)-	764-42-1	0.259	0.015	0.831	3.389	2.34
2-chlorobenzyl chloride	611-19-8	0.293	0.021	0.736	2.965	2.78
2-decanone	693-54-9	0.369	0.079	1.494	7.280	
2-hydroxy-1,4-naphthoquinone	83-72-7	0.842	0.019	0.699	4.215	
2-mercaptobenzothiazole	149-30-4	1.035	0.027	0.577	4.063	
2-mercaptoethanol	60-24-2	-1.785	0.044	1.256	0.293	0.43
2-methyl-1,4-naphthoquinone	58-27-5	0.646	0.014	0.814	4.322	
2-methyl-4-nitroimidazole	696-23-1	-0.062	0.013	0.983	3.365	
2-methylthio-4-*tert*-butylamino-6-amino-*s*-triazine	30125-65-6	0.762	0.028	1.100	6.180	
2-oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	470-82-6	-1.121	0.027	1.316	2.373	
2-phenylindole	948-65-2	1.668	0.031	0.772	6.795	2.85
2-phenylphenol	90-43-7	0.951	0.016	0.800	5.052	
2-propenenitrile, 2-chloro-	920-37-6	-0.502	0.043	0.743	0.901	3.02
2-propenenitrile	107-13-1	-1.196	0.038	1.019	0.564	1.02
2-undecanone	112-12-9	0.687	0.097	1.495	8.125	
2-vinylpyridine	100-69-6	-0.290	0.017	0.955	2.611	1.21
3-(methylthio)propionaldehyde	3268-49-3	-1.355	0.032	1.186	1.049	1.17
3,3'-dichlorobenzidine	91-94-1	2.132	0.040	0.613	7.158	2.70
3,3'-thiodipropionic acid	111-17-1	-0.077	0.013	1.033	3.598	
3,4,5-trichlorocatechol	56961-20-7	1.322	0.046	0.387	3.790	
3,4,6-trichlorocatechol	32139-72-3	1.357	0.045	0.387	3.883	
3,4-dichlorocatechol	3978-67-4	0.907	0.037	0.499	3.302	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
3,5,5-trimethyl-2-cyclohexen-1-one	78-59-1	-0.743	0.020	1.227	2.889	
3,5-dichlorocatechol	13673-92-2	0.971	0.036	0.499	3.471	
3,5-di-*tert*-butylsalicylic acid	19715-19-6	0.600	0.028	1.144	5.990	1.97
3a,4,7,7a-tetrahydro-1H-indene	3048-65-5	-0.572	0.018	1.061	2.440	1.44
3-amino-1,2,4-triazole	61-82-5	-0.690	0.041	0.810	0.766	
3-amino-2-butenenitrile	1118-61-2	-0.613	0.020	1.014	2.077	0.68
3-amino-4-chlorobenzoic acid	2840-28-0	1.117	0.023	0.622	4.524	
3-aminophenol	591-27-5	-0.129	0.026	0.785	2.113	-0.04
3-aminopyridine	462-08-8	-0.576	0.031	0.859	1.333	1.04
4,4,4-trifluorocrotonitrile	406-86-0	-0.495	0.018	1.015	2.394	2.78
4,4'-dihydroxy-biphenyl	92-88-6	1.651	0.030	0.762	6.695	1.16
4,5-dichlorocatechol	3428-24-8	1.014	0.035	0.499	3.583	
4,6-dinitro-*o*-cresol	534-52-1	0.857	0.015	0.835	4.993	2.26
4-amino-2-nitrophenol	119-34-6	0.614	0.019	0.713	3.688	
4-aminophenol	123-30-8	0.065	0.021	0.785	2.627	2.07
4-aminopyridine	504-24-5	-0.569	0.031	0.859	1.353	1.44
4-chlorocatechol	2138-22-9	0.580	0.027	0.622	3.105	
4-ethyl-1,1'-biphenyl	5707-44-8	1.321	0.030	0.931	6.740	2.48
4-hydroxybenzoic acid	99-96-7	0.692	0.017	0.751	4.100	0.17
4-methylbenzoic acid	99-94-5	0.436	0.013	0.901	4.239	0.33
4-vinylpyridine	100-43-6	-0.284	0.016	0.955	2.627	2.02
5-ethylidene-8,9,10-trinorborn-2-ene	16219-75-3	-0.486	0.016	1.061	2.668	1.23
6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	91-53-2	0.831	0.030	1.099	6.356	1.51
6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	2439-01-2	1.888	0.034	0.667	6.807	
8-hydroxyquinoline	148-24-3	0.329	0.019	0.763	3.205	1.76
Acetophenone	98-86-2	-0.127	0.014	0.956	3.048	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (h^*=0.111)	GATS1v	DP07	pLC_{50} (mM)
Acetylsalicylic acid	50-78-2	0.354	0.013	0.909	4.064	
alpha,alpha-dichlorotoluene	98-87-3	0.352	0.020	0.736	3.121	0.85
Anisol	100-66-3	-0.492	0.017	1.037	2.521	
Atrazine-deisopropyl	1007-28-9	0.645	0.014	0.918	4.882	
Benazolin ethyl	25059-80-7	1.183	0.021	0.847	5.921	
Benzaldehyde	100-52-7	-0.170	0.018	0.882	2.533	
Benzenethiol	108-98-5	-0.441	0.026	0.869	1.746	4.09
Benzoyl-chloride	98-88-4	0.331	0.019	0.753	3.157	
Benzyl-chloroformiate	501-53-1	0.426	0.013	0.868	4.031	
Betariboacetate	13035-61-5	0.786	0.050	1.255	7.085	
Biphenyl	92-52-4	0.882	0.016	0.837	5.069	1.60
Bis(2-chloroethyl) ether	111-44-4	-0.682	0.019	1.107	2.399	
Bismerthiazol	79319-85-0	2.073	0.038	0.536	6.584	
Buparvaquone	88426-33-9	1.791	0.065	1.040	8.575	
But-3-en-3-olide	674-82-8	-0.801	0.026	0.991	1.455	0.98
Butachlor	23184-66-9	0.401	0.029	1.199	5.763	3.05
Butylbenzyl phthalate	85-68-7	2.021	0.075	1.018	9.063	
Captopril	62571-86-2	0.159	0.020	1.167	4.948	
Chlorohydroquinone	615-67-8	0.508	0.029	0.622	2.914	
Cinmethylin	87818-31-3	0.656	0.032	1.164	6.247	
Cotinine	486-56-6	0.352	0.016	1.053	4.840	
Cyclohexanone oxime	100-64-1	-0.930	0.022	1.231	2.418	
Cyclohexanone	108-94-1	-1.418	0.032	1.256	1.262	
Cyclosulfamuron	136849-15-5	2.291	0.088	0.993	9.643	
Diallyl phthalate	131-17-9	1.498	0.047	1.024	7.713	2.75
Diethyl disulfide	110-81-6	-0.907	0.022	1.126	1.907	1.01

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Diethyl malonate	105-53-3	-0.079	0.043	1.397	5.570	0.61
Diethyl phthalate	84-66-2	0.698	0.026	1.101	6.016	
Diisobutyl phthalate	84-69-5	0.673	0.031	1.151	6.221	1.92
Dimethyl disulphide	624-92-0	-0.952	0.032	0.984	1.018	1.93
Dimethylformamide	68-12-2	-2.407	0.066	1.627	0.660	
Dimethylnitrosamine	62-75-9	-2.541	0.073	1.677	0.579	
Dinitramine	29091-05-2	0.765	0.027	1.092	6.145	
Diphenylpropanediol		0.890	0.018	0.911	5.492	
Enrofloxacin	93106-60-6	2.066	0.079	1.023	9.211	
Ethanethiol	75-08-1	-2.101	0.052	1.385	0.157	1.45
Flumazenil	78755-81-4	1.611	0.047	0.987	7.813	
Glycidyl methacrylate	106-91-2	0.041	0.022	1.210	4.871	1.71
Hexamethylene diacrylate	13048-33-4	0.206	0.034	1.284	5.708	2.77
Hydrogenatedbisphenol A	80-04-6	0.677	0.055	1.307	7.079	1.10
Hydroquinone	123-31-9	0.052	0.022	0.776	2.544	
Isoprene	78-79-5	-1.712	0.040	1.300	0.723	0.66
Maleic anhydride	108-31-6	-0.507	0.028	0.863	1.537	
Medazepam	2898-12-6	1.310	0.026	0.876	6.414	
Methacrylonitrile	126-98-7	-1.332	0.035	1.126	0.783	
Methyl isothiocyanate	556-61-6	-0.721	0.032	0.900	1.172	2.78
Methylhydrazine	60-34-4	-2.836	0.085	1.726	0.065	2.08
m-toluic acid	99-04-7	0.222	0.013	0.901	3.672	0.22
n-(*tert*-butyl)-2-benzothiazolylsulfenamide	95-31-8	1.511	0.034	0.897	7.060	2.23
n,*n*-dimethylhydrazine	57-14-7	-2.973	0.095	1.806	0.137	
n-cyclohexyl-2-benzothiazolylsulfenamide	95-33-0	1.555	0.034	0.883	7.098	2.10
Nitroglycerin	55-63-0	0.097	0.028	1.252	5.247	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Nitroglycerin	55-63-0	0.097	0.028	1.252	5.247	
n-methyl-*n*,*n*-bis(2-dimethylaminoethyl)amine	3030-47-5	-0.789	0.088	1.740	5.552	
o-acetoacetotoluidide	93-68-5	0.830	0.024	1.033	5.997	
o-chlorobenzonitrile	873-32-5	0.383	0.024	0.689	2.946	0.57
Octanedinitrile	629-40-3	0.733	0.045	1.238	6.852	-0.61
Olaquindox	23696-28-8	1.183	0.026	0.941	6.430	
Ondansetron	99614-02-5	1.346	0.034	0.973	7.034	
o-tolidine	119-93-7	1.589	0.034	0.873	7.134	1.21
Pentane-1-thiol	110-66-7	-1.364	0.036	1.436	2.383	**1.90**
Perfluorooctane sulfonic acid	1763-23-1	1.317	0.103	1.362	9.069	
Phenol,4,4',4''-ethylidynetris-	27955-94-8	1.825	0.039	0.812	7.427	
Phthalic anhydride	85-44-9	0.381	0.019	0.753	3.288	
Phthalonitrile	91-15-6	0.438	0.019	0.734	3.336	0.75
Pivaloyl chloride	3282-30-2	-1.310	0.030	1.205	1.271	
p-methoxybenzaldehyde	123-11-5	0.426	0.014	0.992	4.706	0.53
p-phenylphenol	92-69-3	1.289	0.022	0.800	5.944	1.70
Propyl gallate	121-79-9	1.363	0.033	0.955	6.983	
Propyzamide	23950-58-5	1.563	0.030	0.814	6.747	
Pyrazosulfuron ethyl	93697-74-6	1.687	0.070	1.107	8.664	
Quinoline	91-22-5	-0.265	0.014	1.020	3.030	0.30
Secobarbital	76-73-3	0.562	0.029	1.160	5.977	
Simazine	122-34-9	0.606	0.024	1.098	5.757	
tert-butyl 2-ethylperoxyhexanoate	3006-82-4	-0.224	0.054	1.488	5.679	1.67
tert-butylhydroperoxide	75-91-2	-2.275	0.060	1.564	0.668	-0.02
Tetrachlorocatechol	1198-55-6	1.712	0.061	0.260	4.132	
Tetrachlorohydroquinone	87-87-6	1.718	0.060	0.260	4.147	

Table C1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[GFS]}$ by Eq. 4.7 (mM)	Hat value (*h**=0.111)	GATS1v	DP07	pLC_{50} (mM)
Tetrachlorophthalic anhydride	117-08-8	1.380	0.025	0.601	5.105	
Tetrahydromethylphthalic anhydride	11070-44-3	-0.098	0.013	1.011	3.423	
Thiopental	76-75-5	0.687	0.026	1.106	6.014	
Thiophene	110-02-1	-0.740	0.056	0.732	0.212	0.43
Thiosemicarbazide	79-19-6	-0.267	0.057	0.603	0.760	0.64
Thiourea dioxide	4189-44-0	-0.430	0.037	0.765	1.210	
Thiourea	62-56-6	-0.390	0.071	0.561	0.208	
Triclosan	3380-34-5	2.143	0.041	0.624	7.248	2.64
Trifluralin	1582-09-8	0.897	0.037	1.138	6.743	
Trimethylquinone	935-92-2	-0.100	0.013	1.028	3.511	
Triphenyl phosphate	115-86-6	1.645	0.040	0.911	7.490	2.40
Tris-(2,3-dibromopropyl) phosphate	126-72-7	1.428	0.026	0.817	6.405	2.56
Tris(2-chloroethyl) phosphate	115-96-8	0.430	0.022	1.125	5.438	

*Bold values indicate pLC_{50} data with very weak relationship with the $pEC_{50,NR[GFS]}$ prediction.

APPENDIX D: EXTERNAL SET CHEMICALS FOR $pEC_{50,NR[FHM]}$ MODEL

Table D1. External set chemicals for $pEC_{50,NR[FHM]}$ model.

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
5-fluorocytosine	2022-85-7	-1.399	0.097	0.204	0.106	0.953
5-fluorouracil	51-21-8	-0.637	0.116	0.213	0.208	1.107
Acetaminophen	103-90-2	-0.812	0.248	0.454	0.030	1.052
Isoniazid	54-85-3	-0.869	0.163	0.308	0.147	0.940
Metformin	657-24-9	-2.210	0.189	0.120	0.042	0.804
Metronidazole	443-48-1	-0.821	0.314	0.458	0.068	0.904
Nitrofurazone	59-87-0	-1.082	0.117	0.312	0.066	1.065
Paracetamol	103-90-2	-0.812	0.248	0.454	0.030	1.052
Piperazine	110-85-0	-0.652	0.145	0.375	0.047	1.286
2,6-dichlorobenzonitrile	1194-65-6	-0.394	0.364	0.433	0.197	0.809
Acroleine	107-02-8	-1.862	0.057	0.000	0.068	1.232
Glyphosate	1071-83-6	-0.520	0.149	0.362	0.066	1.341
1,1-dichloroethane	75-34-3	-2.542	0.168	0.000	0.044	0.842
1,3-dibromopropane	109-64-8	-0.264	0.417	0.000	0.407	1.163
1,3-dichloropropane	142-28-9	-1.309	0.093	0.000	0.187	1.202
1,3-dichloropropene	542-75-6	-1.214	0.109	0.000	0.209	1.192
1-chlorobutane	109-69-3	-1.791	0.058	0.000	0.079	1.243
Bromodichloromethane	75-27-4	-2.067	0.090	0.000	0.000	1.326
Dibromochloromethane	124-48-1	-2.057	0.092	0.000	0.000	1.333
Ethyl bromide	74-96-4	-1.962	0.058	0.000	0.099	1.054
1,3-dichlorobenzene	541-73-1	-0.862	0.260	0.385	0.120	0.862

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
2,4-diaminotoluene	95-80-7	-1.006	0.260	0.424	0.040	0.951
2,5-diaminotoluene	95-70-5	-1.131	0.225	0.386	0.040	0.951
2,6-diamino-4-nitrotoluene	59229-75-3	-0.829	0.397	0.508	0.040	0.881
2,6-diaminotoluene	823-40-5	-1.290	0.193	0.344	0.036	0.951
2,6-dinitrotoluene	606-20-2	-0.749	0.337	0.485	0.060	0.920
2-amino-4-nitrotoluene	99-55-8	-0.813	0.335	0.480	0.051	0.918
2-amino-6-nitrotoluene	603-83-8	-0.931	0.266	0.420	0.067	0.918
2-chlorotoluene	95-49-8	-0.743	0.284	0.388	0.150	0.833
2-phenylpropene	98-83-9	-1.092	0.253	0.412	0.031	0.950
3-chlorotoluene	108-41-8	-0.901	0.266	0.372	0.129	0.833
Benzyl alcohol	100-51-6	-1.050	0.122	0.332	0.046	1.111
Benzyl cyanide	140-29-4	-0.887	0.166	0.392	0.034	1.128
Benzylamine	100-46-9	-1.035	0.136	0.350	0.036	1.115
Bromobenzene	108-86-1	0.074	0.351	0.411	0.270	0.931
Catechol	120-80-9	-0.999	0.075	0.259	0.096	1.140
Chlorobenzene	108-90-7	-0.866	0.139	0.318	0.121	1.010
Divinylbenzene	1321-74-0	-0.798	0.286	0.475	0.030	1.013
Ethylbenzene	100-41-4	-0.931	0.192	0.407	0.028	1.084
Isopropylbenzene	98-82-8	-0.767	0.285	0.479	0.030	1.026
m-phenylenediamine	108-45-2	-1.214	0.091	0.278	0.052	1.100
o-phenylenediamine	95-54-5	-1.127	0.094	0.286	0.064	1.100
p-phenylenediamine	106-50-3	-1.107	0.127	0.333	0.037	1.100
Resorcinol	108-46-3	-1.269	0.074	0.253	0.046	1.140
Toluene	108-88-3	-1.379	0.141	0.304	0.026	1.016
Trans-cinnamic acid	140-10-3	-0.528	0.222	0.459	0.053	1.158
2,3-dimethylphenol	526-75-0	-1.284	0.263	0.370	0.045	0.864

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
2,4,6-trimethylphenol	527-60-6	-1.131	0.408	0.454	0.043	0.784
2,6-dimethylphenol	576-26-1	-1.294	0.260	0.367	0.045	0.864
2-allylphenol	1745-81-9	-0.564	0.255	0.482	0.043	1.115
2-ethylphenol	90-00-6	-0.833	0.200	0.413	0.051	1.058
2-methoxyphenol	90-05-1	-1.298	0.214	0.329	0.066	0.875
3,4-dimethylphenol	95-65-8	-1.138	0.325	0.431	0.034	0.864
3-ethylphenol	620-17-7	-0.716	0.237	0.450	0.050	1.058
4-(1-methylethenyl)phenol	4286-23-1	-1.034	0.293	0.437	0.034	0.922
4-chloro-2-methylphenol	1570-64-5	-0.771	0.365	0.448	0.114	0.801
4-ethylphenol	123-07-9	-0.788	0.237	0.448	0.037	1.058
Methyl *p*-hydroxybenzoate	99-76-3	-0.927	0.332	0.431	0.087	0.826
2-(2-butoxyethoxy)ethanol	112-34-5	-0.747	0.121	0.345	0.056	1.257
2-butoxyethanol	111-76-2	-1.088	0.067	0.220	0.056	1.307
2-isoproxyethanol	109-59-1	-1.540	0.036	0.105	0.080	1.173
3-pentanol	71-41-0	-1.796	0.081	0.000	0.051	1.337
Hexanol	111-27-3	-0.845	0.120	0.256	0.042	1.442
1,6-hexanediamine	124-09-4	-0.633	0.192	0.329	0.011	1.530
2-amino-2-ethylpropanediol	115-70-8	-1.761	0.059	0.000	0.085	1.243
Dibutylamine	111-92-2	-0.478	0.202	0.427	0.021	1.378
Diethanolamine	111-42-2	-1.285	0.089	0.112	0.056	1.418
Diethylamine	109-89-7	-2.034	0.069	0.000	0.025	1.262
Diethylnitrosamine	55-18-5	-2.017	0.055	0.000	0.076	1.096
Diisopropylamine	108-18-9	-2.193	0.069	0.000	0.018	1.176
Dimethylamine	124-40-3	-2.774	0.224	0.000	0.006	0.813
t-butylamine	75-64-9	-2.473	0.108	0.000	0.007	1.019
Triethylamine	121-44-8	-2.257	0.073	0.000	0.028	1.096

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
2-(dimethylamino)ethyl methacrylate	2867-47-2	-1.911	0.257	0.220	0.053	0.744
2-hydroxyethyl acrylate	818-61-1	-1.716	0.039	0.087	0.069	1.130
2-hydroxyethyl methacrylate	868-77-9	-1.784	0.079	0.129	0.075	0.965
Ethylacrylate	140-88-5	-1.929	0.058	0.059	0.075	1.025
Isobutyl acetate	110-19-0	-1.477	0.093	0.235	0.049	1.026
Methyl acrylate	96-33-3	-2.498	0.207	0.000	0.078	0.754
Methyl methacrylate	80-62-6	-2.533	0.285	0.000	0.102	0.646
n-butyl acrylate	141-32-2	-1.333	0.046	0.185	0.066	1.182
n-butyl methacrylate	97-88-1	-1.480	0.079	0.225	0.045	1.061
Vinyl acetate	108-05-4	-2.275	0.098	0.000	0.067	0.948
2,3-dimethylaniline	87-59-2	-1.283	0.273	0.382	0.035	0.872
2,4-dimethylaniline	95-68-1	-1.004	0.346	0.459	0.040	0.872
2,6-dimethylaniline	87-62-7	-1.298	0.271	0.379	0.034	0.872
2-ethylaniline	578-54-1	-0.845	0.216	0.427	0.040	1.056
2-methyl-4-nitroaniline	99-52-5	-0.836	0.327	0.473	0.051	0.918
2-methylaniline	95-53-4	-1.262	0.165	0.330	0.042	0.982
2-nitroaniline	88-74-4	-1.003	0.128	0.338	0.056	1.095
2-nitro-*p*-anisidine	96-96-8	-1.002	0.414	0.479	0.045	0.810
3,4-dimethylaniline	95-64-7	-1.049	0.355	0.462	0.029	0.872
3,5-dichloroaniline	626-43-7	-1.031	0.314	0.381	0.113	0.778
3,5-dimethylaniline	108-69-0	-1.444	0.265	0.359	0.018	0.872
3-chloroaniline	108-42-9	-0.898	0.175	0.331	0.126	0.940
3-ethylaniline	587-02-0	-0.698	0.268	0.476	0.037	1.056
3-methylaniline	108-44-1	-1.326	0.159	0.318	0.037	0.982
3-nitroaniline	99-09-2	-0.942	0.145	0.361	0.053	1.095
4-chloroaniline	106-47-8	-0.643	0.271	0.440	0.105	0.940

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
4-ethylaniline	589-16-2	-0.748	0.271	0.476	0.027	1.056
4-fluoroaniline	371-40-4	-1.099	0.100	0.295	0.066	1.092
4-methylaniline	106-49-0	-1.056	0.241	0.415	0.027	0.982
4-nitroaniline	100-01-6	-0.890	0.151	0.369	0.058	1.095
n,n'-dimethylaniline		-1.361	0.374	0.414	0.021	0.793
n,n-Dimethylaniline	121-69-7	-1.361	0.374	0.414	0.021	0.793
n-Ethylaniline	103-69-5	-0.783	0.207	0.434	0.030	1.118
p-anisidine	104-94-9	-1.333	0.258	0.369	0.031	0.881
2-ethyl butanoic acid	88-09-5	-1.938	0.059	0.000	0.052	1.235
2-methyl butanoic acid	600-07-7	-1.906	0.052	0.000	0.075	1.177
3-methyl butanoic acid	503-74-2	-2.106	0.061	0.000	0.035	1.177
Acrylic acid	79-10-7	-1.813	0.053	0.000	0.087	1.200
Adipic acid	124-04-9	-0.807	0.114	0.276	0.047	1.405
A-fluoro-b-alanine	3821-81-6	-0.849	0.195	0.000	0.254	1.289
Amidosulfonic acid	5329-14-6	-2.123	0.065	0.000	0.025	1.200
Gentisic acid	490-79-9	-0.957	0.201	0.397	0.049	1.016
Glyoxylic acid	298-12-4	-0.293	0.400	0.000	0.395	1.185
Heptanoic acid	111-14-8	-0.619	0.162	0.372	0.026	1.389
Isocyanuric acid	108-80-5	-2.283	0.270	0.134	0.046	0.707
Malonic acid diethylester	105-53-3	-1.579	0.072	0.185	0.061	1.028
Methacrylic acid	79-41-4	-2.369	0.132	0.000	0.069	0.875
Octanoic acid	124-07-2	-0.179	0.278	0.497	0.027	1.404
Perfluorooctanoic acid	335-67-1	0.106	0.263	0.452	0.163	1.232
Pivalic acid	75-98-9	-2.388	0.146	0.000	0.073	0.848
Sorbic acid	110-44-1	-1.765	0.044	0.103	0.028	1.202
2-methylvaleraldehyde	123-15-9	-1.557	0.065	0.083	0.041	1.348

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h*=0.429)	X5v	R3m+	GATS2v
Capronaldehyde	66-25-1	-1.113	0.079	0.203	0.043	1.374
Crotonaldehyde	4170-30-3	-2.140	0.063	0.000	0.032	1.164
Glutaraldehyde	111-30-8	-1.379	0.077	0.096	0.057	1.386
Propionaldehyde	123-38-6	-1.898	0.058	0.000	0.060	1.235
2-chloronitrobenzene	88-73-3	-0.186	0.255	0.396	0.209	0.997
2-nitroanisole	91-23-6	-1.044	0.325	0.431	0.057	0.849
2-nitrotoluene	88-72-2	-1.034	0.210	0.382	0.059	0.962
3-chloronitrobenzene	121-73-3	-0.463	0.255	0.444	0.122	0.997
3-nitroanisole	555-03-3	-1.058	0.333	0.436	0.051	0.849
4-amino-2-nitrotoluene	119-32-4	-0.772	0.365	0.503	0.044	0.918
4-chloronitrobenzene	100-00-5	-0.562	0.271	0.464	0.089	0.997
4-methylnitrobenzene	99-99-0	-0.869	0.267	0.441	0.053	0.962
Nitrobenzene	98-95-3	-0.964	0.099	0.312	0.055	1.185
Styrene-7,8-oxide	96-09-3	-0.545	0.273	0.492	0.048	1.088
1,1'-oxybis-butane	142-96-1	-0.805	0.129	0.348	0.027	1.311
1,2,3-trihydroxybenzene	87-66-1	-1.007	0.087	0.268	0.103	1.089
1-methoxy-2-propanol	107-98-2	-2.160	0.073	0.000	0.071	1.014
2-aminophenol	95-55-6	-0.750	0.100	0.273	0.144	1.114
2-aminopyridine	504-29-0	-1.341	0.058	0.206	0.074	1.100
2-butanone oxime	96-29-7	-2.047	0.056	0.000	0.066	1.110
2-butenedinitrile, (E)-	764-42-1	-1.709	0.091	0.017	0.044	1.383
2-chlorobenzyl chloride	611-19-8	-0.089	0.356	0.501	0.170	0.959
2-mercaptoethanol	60-24-2	-0.507	0.308	0.000	0.312	1.325
2-methyl-4-nitroimidazole	696-23-1	-1.600	0.125	0.212	0.072	0.913
2-propenenitrile	107-13-1	-1.969	0.052	0.000	0.069	1.154
2-vinylpyridine	100-69-6	-1.345	0.104	0.276	0.038	1.062

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h^*=0.429)	X5v	R3m+	GATS2v
3-(methylthio)propionaldehyde	3268-49-3	-0.807	0.097	0.144	0.162	1.308
3-amino-1,2,4-triazole	61-82-5	-2.048	0.064	0.033	0.069	1.023
3-amino-2-Butenenitrile	1118-61-2	-2.253	0.084	0.000	0.057	0.998
3-aminophenol	591-27-5	-1.151	0.080	0.265	0.069	1.114
3-aminopyridine	462-08-8	-1.291	0.065	0.199	0.103	1.050
4,4,4-trifluorocrotonitrile	406-86-0	-1.562	0.046	0.042	0.108	1.205
4-amino-2-nitrophenol	119-34-6	-0.780	0.193	0.404	0.071	1.046
4-aminophenol	123-30-8	-1.060	0.104	0.308	0.059	1.114
4-aminopyridine	504-24-5	-1.401	0.066	0.199	0.081	1.050
4-chlorocatechol	2138-22-9	-0.780	0.237	0.398	0.110	0.924
4-hydroxybenzoic acid	99-96-7	-0.843	0.155	0.366	0.078	1.065
4-methylbenzoic acid	99-94-5	-0.770	0.309	0.466	0.066	0.928
4-vinylpyridine	100-43-6	-1.508	0.119	0.260	0.033	1.003
Acetophenone	98-86-2	-1.038	0.204	0.384	0.051	0.982
alpha,alpha-dichlorotoluene	98-87-3	0.050	0.344	0.509	0.171	1.034
Anisol	100-66-3	-1.462	0.198	0.316	0.031	0.913
Atrazine-deisopropyl	1007-28-9	-1.061	0.261	0.360	0.107	0.826
Benzaldehyde	100-52-7	-1.127	0.104	0.302	0.053	1.102
Benzenethiol	108-98-5	-0.790	0.161	0.342	0.124	0.997
Benzoyl-chloride	98-88-4	-0.629	0.200	0.396	0.116	1.013
Benzyl-chloroformiate	501-53-1	-0.030	0.343	0.503	0.168	1.003
Bis(2-chloroethyl) ether	111-44-4	-0.478	0.143	0.231	0.228	1.107
But-3-en-3-olide	674-82-8	-2.313	0.085	0.000	0.037	1.026
Chlorohydroquinone	615-67-8	0.034	0.338	0.355	0.301	0.924
Diethyl malonate	105-53-3	-1.579	0.072	0.185	0.061	1.028
Dimethyl disulphide	624-92-0	-2.412	0.104	0.000	0.028	0.988

Table D1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ NR[FHM]}$ by Eq. 4.10 (mM)	Hat value (h*=0.429)	X5v	R3m+	GATS2v
Ethanethiol	75-08-1	-2.117	0.063	0.000	0.031	1.183
Glycidyl methacrylate	106-91-2	-1.571	0.105	0.204	0.073	0.948
Hydroquinone	123-31-9	-1.190	0.091	0.286	0.040	1.140
Isoprene	78-79-5	-2.346	0.098	0.000	0.045	0.975
Maleic anhydride	108-31-6	-1.825	0.048	0.069	0.080	1.057
Methacrylonitrile	126-98-7	-2.442	0.186	0.000	0.082	0.779
Methylhydrazine	60-34-4	-2.493	0.117	0.000	0.013	0.984
m-toluic acid	99-04-7	-0.848	0.293	0.450	0.061	0.928
Nitroglycerin	55-63-0	-1.003	0.074	0.256	0.059	1.273
o-chlorobenzonitrile	873-32-5	-0.374	0.251	0.376	0.204	0.930
Octanedinitrile	629-40-3	-0.427	0.189	0.414	0.046	1.356
Pentane-1-thiol	110-66-7	-0.587	0.128	0.335	0.076	1.322
Phthalonitrile	91-15-6	-1.095	0.179	0.363	0.048	1.001
Pivaloyl chloride	3282-30-2	-2.091	0.226	0.000	0.178	0.689
p-methoxybenzaldehyde	123-11-5	-1.183	0.322	0.408	0.051	0.826
tert-butylhydroperoxide	75-91-2	-2.239	0.096	0.000	0.075	0.945
Thiophene	110-02-1	-2.246	0.072	0.000	0.018	1.139
Thiosemicarbazide	79-19-6	-1.345	0.154	0.000	0.262	0.916
Thiourea dioxide	4189-44-0	-1.894	0.074	0.000	0.133	0.983
Thiourea	62-56-6	-2.399	0.104	0.000	0.033	0.980

APPENDIX E: EXTERNAL SET CHEMICALS FOR $pEC_{50,MTT[PLHC-1]}$ MODEL

Table E1. External set chemicals for $pEC_{50,MTT[PLHC-1]}$ model.

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h^*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
5-fluorocytosine	2022-85-7	-1.118	0.205	0.247	0	0	
5-fluorouracil	51-21-8	-1.031	0.189	0.183	0	0	
Acebutolol	37517-30-9	0.844	0.072	-0.526	3	0	
Acetaminophen	103-90-2	-0.015	0.058	-0.119	2	0	
Acridine	260-94-6	-0.324	0.137	-0.333	0	0	
Acrivastine	87848-99-5	0.217	0.101	-0.475	4	1	
Acyclovir	59277-89-3	-0.205	0.070	0.020	2	0	
Alprenolol	13655-52-2	0.448	0.070	-0.457	2	0	
Amitriptyline	50-48-6	0.890	0.058	-0.340	4	0	
Amobarbital	57-43-2	0.153	0.178	-0.681	0	0	
Ampicillin	69-53-4	-1.152	0.121	-0.135	1	1	
Antipyrine	60-80-0	-0.071	0.087	-0.298	1	0	
Atovaquone	95233-18-4	0.494	0.054	-0.051	4	0	
Bumetanide	28395-03-1	-0.264	0.097	0.096	5	1	
Bupropion	34841-39-9	0.814	0.120	-0.724	2	0	
Caffeine	58-08-2	-0.029	0.059	-0.109	2	0	
Camptothecin	7689-03-4	0.319	0.061	-0.363	2	0	
Capecitabine	154361-50-9	0.068	0.166	-0.619	0	0	
Carvedilol	72956-09-3	0.664	0.054	-0.395	3	0	
Cefuroxime axetil	64544-07-6	0.433	0.131	-0.666	1	0	
Cefuroxime	55268-75-2	-1.518	0.147	0.132	1	1	
Cephalexin	15686-71-2	-1.319	0.128	-0.013	1	1	
Chloramphenicol	56-75-7	-0.223	0.072	0.033	2	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h^*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Chlorotetracycline	57-62-5	0.574	0.051	-0.109	4	0	
Chlorpromazine	50-53-3	1.376	0.104	-0.475	5	0	
Chlorpropamide	94-20-2	-0.201	0.070	0.017	2	0	
Cimetidine	51481-61-9	0.358	0.063	-0.391	2	0	
Clonazepam	1622-61-3	0.428	0.107	0.217	5	0	
Clonidine	4205-90-7	0.129	0.053	-0.004	3	0	
Cromolyn	16110-51-3	-0.583	0.313	-0.078	6	2	
Cytarabine	147-94-4	-0.650	0.144	-0.095	0	0	
Deoxytetracycline	564-25-0	0.077	0.057	0.034	3	0	
Desipramine	50-47-5	0.756	0.052	-0.242	4	0	
Doxycycline	564-25-0	0.751	0.062	-0.458	3	0	
Estradiol	50-28-2	0.489	0.045	-0.267	3	0	
Ethinylestradiol	57-63-6	0.547	0.047	-0.309	3	0	
Famciclovir	104227-87-4	0.117	0.096	-0.435	1	0	
Famotidine	76824-35-6	-0.215	0.071	0.027	2	0	
Fleroxacin	79660-72-3	0.200	0.095	-0.243	5	1	
Florfenicol	73231-34-2	-0.015	0.058	-0.119	2	0	
Flunitrazepam	1622-62-4	0.323	0.123	0.294	5	0	
Flutamide	13311-84-7	0.347	0.044	-0.163	3	0	
Furaltadone	139-91-3	0.065	0.056	-0.177	2	0	
Gemcitabine	95058-81-4	-0.851	0.163	0.052	0	0	
Glibenclamid	10238-21-8	1.823	0.187	-0.801	5	0	
Hydrochlorothiazide	58-93-5	-0.217	0.152	0.468	4	0	
Hydrocortisone	50-23-7	-0.610	0.142	-0.124	0	0	
Imipramine	50-49-7	0.890	0.058	-0.340	4	0	
Indomethacin	53-86-1	-0.044	0.089	-0.065	5	1	
Isoniazid	54-85-3	-0.253	0.088	-0.165	1	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h^*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Isotretinoin	4759-48-2	-0.560	0.254	-0.787	0	1	
Ketoconazole	65277-42-1	1.275	0.093	-0.401	5	0	
Ketoprofen	22071-15-4	-0.662	0.069	-0.053	3	1	
Labetalol	36894-69-6	0.719	0.051	-0.215	4	0	
Lamotrigine	84057-84-1	0.799	0.139	0.166	6	0	
Lansoprazole	103577-45-3	0.661	0.050	-0.173	4	0	
Levofloxacin	100986-85-4	0.250	0.098	-0.279	5	1	
Lidocaine	137-58-6	0.956	0.087	-0.608	3	0	
Lomefloxacin	98079-51-7	0.250	0.098	-0.279	5	1	
L-tryptophan	73-22-3	-1.440	0.138	0.075	1	1	
Metaxalone	1665-48-1	0.430	0.044	-0.224	3	0	
Metformin	657-24-9	-0.981	0.181	0.147	0	0	
Methotrexate	59-05-2	-0.295	0.333	-0.508	5	2	
Methylprednisolone	83-43-2	0.313	0.206	-0.798	0	0	
Metipranolol	22664-55-7	1.343	0.100	-0.451	5	0	
Metoprolol	37350-58-6	0.267	0.059	-0.325	2	0	
Metronidazole	443-48-1	0.137	0.056	-0.230	2	0	
Midazolam	59467-70-8	1.358	0.183	-0.022	7	0	
Nadolol	42200-33-9	0.789	0.066	-0.486	3	0	
Naproxen	22204-53-1	-0.591	0.090	-0.325	2	1	
Nicotine	22083-74-5	0.109	0.095	-0.429	1	0	
Nisoldipine	63675-72-9	0.524	0.078	-0.512	2	0	
Nitrofurazone	59-87-0	-0.241	0.074	0.046	2	0	
Norfloxacin	70458-96-7	0.003	0.079	-0.319	4	1	
Ofloxacin	82419-36-1	0.241	0.097	-0.273	5	1	
Oxprenolol	6452-71-7	0.419	0.068	-0.436	2	0	
Oxyphenbutazone	129-20-4	0.481	0.045	-0.261	3	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h^*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Oxytetracycline	79-57-2	0.440	0.044	-0.231	3	0	
Paracetamol	103-90-2	-0.016	0.059	-0.118	2	0	
Paroxetine	61869-08-7	1.394	0.106	-0.488	5	0	
Pentobarbital	76-74-4	-0.065	0.152	-0.522	0	0	
Phenobarbital	50-06-6	-0.252	0.088	-0.166	1	0	
Phenoxymethylpenicillinic Acid	87-08-1	-1.038	0.120	-0.218	1	1	
Phenylbutazone	50-33-9	0.067	0.056	-0.179	2	0	
Phenytoin	57-41-0	0.104	0.056	-0.206	2	0	
Pindolol	13523-86-9	-0.035	0.088	-0.324	1	0	
Piperazine	110-85-0	-0.403	0.136	-0.275	0	0	
Prazosin	19216-56-9	1.756	0.171	-0.752	5	0	
Prednisolone	50-24-8	-0.750	0.152	-0.022	0	0	
Procaine	59-46-1	0.302	0.060	-0.350	2	0	
Progesterone	57-83-0	-0.172	0.144	-0.444	0	0	
Promazine	58-40-2	0.793	0.053	-0.269	4	0	
Propylthiouracil	51-52-5	-0.440	0.136	-0.248	0	0	
Quinine	130-95-0	1.099	0.184	-0.932	2	0	
Ranitidine	66357-35-5	0.695	0.100	-0.637	2	0	
Risperidone	106266-06-2	0.607	0.088	-0.573	2	0	
Sotalol	3930-20-9	0.224	0.057	-0.293	2	0	
Sulfaguanidine	57-67-0	-0.579	0.122	0.293	2	0	
Sulfaphenazole	526-08-9	0.787	0.053	-0.265	4	0	
Sulfasalazine	599-79-1	-0.186	0.154	0.259	6	1	
Sumatriptan	103628-46-2	0.293	0.060	-0.344	2	0	
Terazosin	63590-64-7	1.437	0.134	-0.739	4	0	
Terbinafine	91161-71-6	0.439	0.132	-0.670	1	0	
Testosterone	58-22-0	0.013	0.160	-0.579	0	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h^*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Theophylline	58-55-9	-0.271	0.077	0.068	2	0	
Timolol	26839-75-8	0.759	0.111	-0.684	2	0	
Tolazamide	1156-19-0	0.070	0.056	-0.181	2	0	
Tolbutamide	64-77-7	0.030	0.057	-0.152	2	0	
Trenbolone acetate	10161-34-9	0.097	0.170	-0.640	0	0	
Triamterene	396-01-0	0.824	0.076	-0.072	5	0	
Trifluprомazine	146-54-3	1.373	0.104	-0.473	5	0	
Warfarin	81-81-2	0.253	0.046	-0.095	3	0	
2,6-dichlorobenzonitrile	1194-65-6	0.077	0.057	0.034	3	0	
Acetochlor	34256-82-1	0.662	0.054	-0.393	3	0	
Acroleine	107-02-8	-0.696	0.148	-0.061	0	0	
Ametryn	834-12-8	0.823	0.069	-0.511	3	0	
Anilofos	64249-01-0	0.359	0.063	-0.392	2	0	
Azocyclotin	41083-11-8	0.475	0.241	-0.916	0	0	
Benalaxyl	71626-11-4	1.276	0.103	-0.622	4	0	
Benomyl	17804-35-2	-0.168	0.087	-0.227	1	0	
Bensulfuron-methyl	83055-99-6	1.315	0.097	-0.430	5	0	
Beta-cyfluthrin	68359-37-5	0.708	0.051	-0.207	4	0	
Bromoxynil	1689-84-5	0.338	0.063	0.063	4	0	
Chloridazon	1698-60-8	-0.483	0.102	0.003	1	0	
Chlorimuron-ethyl	90982-32-4	0.776	0.078	-0.037	5	0	
Chlorothalonil	1897-45-6	0.661	0.155	0.267	6	0	
Chlorotoluron	15545-48-9	0.369	0.044	-0.179	3	0	
Chlorsulfuron	64902-72-3	1.009	0.077	-0.207	5	0	
Cyanazine	21725-46-2	0.863	0.074	-0.540	3	0	
Cyhexatin	13121-70-5	0.288	0.201	-0.780	0	0	
Decanoic acid	334-48-5	-1.262	0.175	-0.275	0	1	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Diclofop methyl	51338-27-3	1.109	0.080	-0.280	5	0	
Diclofop P	40843-25-2	-0.023	0.089	-0.080	5	1	
Dimethachlon	24096-53-5	-0.042	0.069	0.121	3	0	
Endosulfan	115-29-7	-0.595	0.141	-0.135	0	0	
Ethalfluraline	55283-68-6	0.393	0.059	0.023	4	0	
Ethametsulfuron	111353-84-5	-0.012	0.089	-0.088	5	1	
Ethoxyquin	91-53-2	0.841	0.072	-0.524	3	0	1.51
Fenamiphos	22224-92-6	1.090	0.111	-0.706	3	0	
Fenhexamid	126833-17-8	0.861	0.056	-0.319	4	0	
Fenoxaprop	95617-09-7	-0.411	0.071	-0.017	4	1	
Fenpropidin	67306-00-7	1.084	0.180	-0.921	2	0	
Fentin hydroxide	76-87-9	0.408	0.044	-0.208	3	0	
Fluazifop P	83066-88-0	0.202	0.099	-0.464	4	1	
Flumetsulam	98967-40-9	0.947	0.076	-0.162	5	0	
Fluroxypyr	69377-81-7	-0.283	0.099	0.110	5	1	
Iprodione	36734-19-7	-0.032	0.067	0.113	3	0	
Cybutryne	28159-98-0	1.044	0.102	-0.672	3	0	
Isoproturon	34123-59-6	0.380	0.065	-0.407	2	0	
Lindane	58-89-9	-0.381	0.136	-0.291	0	0	
Mefenacet	73250-68-7	0.554	0.081	-0.534	2	0	
Metalaxyl	57837-19-1	0.834	0.071	-0.519	3	0	
Methabenzthiazuron	18691-97-9	-0.220	0.088	-0.189	1	0	
Metolachlor	51218-45-2	0.912	0.081	-0.576	3	0	
Metribuzin	21087-64-9	-0.029	0.155	-0.548	0	0	
Metsulfuron-methyl	74223-64-6	1.131	0.082	-0.296	5	0	
Molinate	2212-67-1	-0.390	0.136	-0.285	0	0	
Nicosulfuron	111991-09-4	1.327	0.098	-0.439	5	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Oxadiargyl	39807-15-3	0.764	0.052	-0.248	4	0	
Oxadiazon	19666-30-9	1.071	0.074	-0.472	4	0	
Oxadixyl	77732-09-3	0.689	0.056	-0.413	3	0	
Pendimethalin	40487-42-1	0.878	0.076	-0.111	5	0	
Phenmedipham	13684-63-4	0.897	0.058	-0.345	4	0	
Pretilachlor	51218-49-6	1.092	0.111	-0.707	3	0	2.11
Prometryn	7287-19-6	0.997	0.094	-0.638	3	0	
Quinclorac	84087-01-4	-0.765	0.074	0.022	3	1	
Quizalofop P	94051-08-8	-0.067	0.074	-0.268	4	1	
Tebuthiuron	34014-18-1	0.839	0.125	-0.742	2	0	
Terbumeton	33693-04-8	1.031	0.100	-0.663	3	0	
Terbutryn	886-50-0	0.934	0.084	-0.592	3	0	
Thiobencarb	28249-77-6	0.588	0.085	-0.559	2	0	2.30
Thiophanate methyl	23564-05-8	0.159	0.056	-0.246	2	0	
Tribenuron	106040-48-6	0.139	0.092	-0.198	5	1	
1,1,2,2-tetrachloroethane	79-34-5	-0.657	0.145	-0.090	0	0	
1,1,2-trichloroethane	79-00-5	-0.644	0.144	-0.099	0	0	
1,1-dichloroethane	75-34-3	-0.597	0.141	-0.134	0	0	
1,1-dichloroethylene	75-35-4	-0.725	0.150	-0.040	0	0	0.33
1,2,3-trichloropropane	96-18-4	-0.501	0.137	-0.204	0	0	
1,2-dibromo-3-chloropropane	96-12-8	-0.572	0.140	-0.152	0	0	0.78
1,2-dichloroethane	107-06-2	-0.660	0.145	-0.088	0	0	
1,2-dichloropropane	78-87-5	-0.564	0.140	-0.158	0	0	-0.15
1,3-dibromopropane	109-64-8	-0.551	0.139	-0.167	0	0	**1.41**
1,3-dichloropropane	142-28-9	-0.580	0.140	-0.146	0	0	
1,3-dichloropropene	542-75-6	-0.754	0.153	-0.019	0	0	**1.87**
1-chlorobutane	109-69-3	-0.561	0.139	-0.160	0	0	
3,4-dichlorobut-1-ene	760-23-6	-0.562	0.140	-0.159	0	0	0.67

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC\text{-}1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Acetic acid, bromo-, 2-butene-1,4-diyl ester	20679-58-7	-0.432	0.136	-0.254	0	0	
Bromodichloromethane	75-27-4	-0.836	0.161	0.041	0	0	0.77
Carbon tetrachloride	56-23-5	-0.949	0.176	0.123	0	0	**1.31**
Cyclohexane	110-82-7	-0.316	0.137	-0.339	0	0	
Dibromochloromethane	124-48-1	-0.783	0.155	0.002	0	0	0.42
Ethyl bromide	74-96-4	-0.581	0.140	-0.145	0	0	
Ethylcyclohexane	1678-91-7	-0.268	0.139	-0.374	0	0	**2.18**
Methylcyclohexane	108-87-2	-0.132	0.146	-0.473	0	0	**1.67**
Pentachloroethane	76-01-7	-0.739	0.151	-0.030	0	0	
Trichloroethylene	79-01-6	-0.877	0.166	0.071	0	0	0.54
1-(*n*-phenylamino)-naphthalene	90-30-2	-0.086	0.062	-0.067	2	0	**2.50**
1,2-dimethylnaphthalene	573-98-8	0.307	0.060	-0.354	2	0	1.83
1,3-dimethylnapthalene	575-41-7	0.110	0.056	-0.210	2	0	1.94
1,5-napthalenediamine	2243-62-1	-0.701	0.147	0.382	2	0	0.97
1,8-naphthylenediamine	479-27-6	-0.701	0.147	0.382	2	0	1.45
1-methylnaphthalene	90-12-0	-0.282	0.089	-0.144	1	0	**1.40**
2,7-dimethylnaphthalene	582-16-1	0.319	0.061	-0.363	2	0	1.94
2-methylnaphthalene	91-57-6	-0.046	0.088	-0.316	1	0	1.87
3-hydroxy-2-naphthoic acid	92-70-6	-0.969	0.088	-0.049	2	1	
Isopropylnaphthalene	29253-36-9	0.012	0.090	-0.358	1	0	**2.36**
n-phenyl-2-naphthylamine	135-88-6	0.082	0.056	-0.190	2	0	
ß-naphthol	135-19-3	-0.254	0.088	-0.164	1	0	1.56
ß-naphthylamine	91-59-8	-0.419	0.097	-0.044	1	0	1.56
Tetralin	119-64-2	0.035	0.057	-0.155	2	0	1.34
1,2,3-trichlorobenzene	87-61-6	-0.019	0.066	0.104	3	0	1.75
1,2-dichlorobenzene	95-50-1	-0.186	0.069	0.006	2	0	1.59
1,3-dichlorobenzene	541-73-1	-0.197	0.070	0.014	2	0	1.41

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC\text{-}1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
1,4-benzenediamine, *n*,*n*'-bis(1-methylpropyl)-	101-96-2	0.469	0.072	-0.472	2	0	**2.78**
1,4-dichlorobenzene	106-46-7	-0.189	0.069	0.008	2	0	**1.82**
2,4,6-trichlorophenylhydrazine	5329-12-4	0.263	0.071	0.118	4	0	2.33
2,4-diamino-6-nitrotoluene	6629-29-4	-0.093	0.125	0.378	4	0	
2,4-diaminotoluene	95-80-7	-0.195	0.089	0.232	3	0	
2,4-dichlorotoluene	95-73-8	0.151	0.051	-0.020	3	0	
2,4-dinitrotoluene	121-14-2	-0.227	0.095	0.256	3	0	
2,5-diaminotoluene	95-70-5	-0.268	0.102	0.286	3	0	**2.81**
2,5-dichlorotoluene	19398-61-9	0.112	0.054	0.008	3	0	1.60
2,6-diamino-4-nitrotoluene	59229-75-3	0.014	0.105	0.300	4	0	
2,6-diaminotoluene	823-40-5	-0.174	0.086	0.217	3	0	
2,6-dichlorotoluene	118-69-4	0.112	0.054	0.008	3	0	1.85
2,6-dinitrotoluene	606-20-2	-0.366	0.120	0.357	3	0	**0.73**
2-amino-4,6-dinitrotoluene	35572-78-2	-0.025	0.112	0.328	4	0	
2-amino-4-nitrotoluene	99-55-8	-0.103	0.076	0.165	3	0	
2-amino-6-nitrotoluene	603-83-8	-0.208	0.091	0.242	3	0	
2-chlorohydroquinonedimethylether	2100-42-7	0.593	0.049	-0.343	3	0	0.79
2-chlorotoluene	95-49-8	-0.056	0.060	-0.089	2	0	1.22
2-methyl-4-chlorophenoxyacetic acid	94-74-6	-0.573	0.067	-0.118	3	1	
2-phenylpropene	98-83-9	0.101	0.095	-0.423	1	0	1.21
3,4-dichlorotoluene	95-75-0	0.245	0.046	-0.089	3	0	1.57
3,5-bis(trifluoromethyl)benzylamine	85068-29-7	0.406	0.044	-0.206	3	0	1.34
3-chlorotoluene	108-41-8	0.009	0.058	-0.136	2	0	
4,6-dichlororesorcinol	137-19-9	0.271	0.070	0.112	4	0	
4-allyl-1,2-dimethoxybenzene	93-15-2	0.449	0.044	-0.238	3	0	1.10
4-chlororesorcinol	95-88-5	0.137	0.052	-0.010	3	0	
4-chlorotoluene	106-43-4	0.085	0.056	-0.192	2	0	1.32

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
4-nitrotoluene-2-sulphonic acid	121-03-9	-1.874	0.287	0.205	3	2	
4-*tert*-butyltoluene	98-51-1	0.645	0.093	-0.601	2	0	1.94
4-toluenesulfonyl chloride	98-59-9	0.018	0.057	-0.143	2	0	0.39
Atenolol	29122-68-7	0.335	0.062	-0.374	2	0	
Benzalacetone	122-57-6	-0.291	0.090	-0.137	1	0	1.41
Benzotrifluoride	98-08-8	-0.405	0.096	-0.054	1	0	1.01
Benzyl alcohol	100-51-6	-0.298	0.090	-0.132	1	0	
Benzyl cyanide	140-29-4	-0.231	0.088	-0.181	1	0	
Benzylamine	100-46-9	-0.213	0.087	-0.194	1	0	
Bromobenzene	108-86-1	-0.342	0.092	-0.100	1	0	1.56
Butylbenzene	104-51-8	-0.086	0.087	-0.287	1	0	1.61
Catechol	120-80-9	0.058	0.056	-0.172	2	0	
Chlorobenzene	108-90-7	-0.378	0.094	-0.074	1	0	1.23
Cyclohexylbenzene	827-52-1	0.083	0.093	-0.410	1	0	**2.13**
Diisopropylbenzene	25321-09-9	0.470	0.072	-0.473	2	0	2.30
Dimethyl phthalate	131-11-3	0.062	0.056	-0.175	2	0	
Diuron	330-54-1	0.223	0.047	-0.073	3	0	
Divinylbenzene	1321-74-0	0.136	0.056	-0.229	2	0	1.49
Ethylbenzene	100-41-4	-0.102	0.087	-0.275	1	0	
Fenobucarb	3766-81-2	0.332	0.062	-0.372	2	0	1.33
Isopropylbenzene	98-82-8	-0.068	0.087	-0.300	1	0	
Metaxylene hexafluoride	402-31-3	-0.335	0.084	0.115	2	0	1.47
m-phenylenebis(methylamine)	1477-55-0	0.287	0.060	-0.339	2	0	0.19
m-phenylenediamine	108-45-2	-0.548	0.117	0.270	2	0	
n-propylbenzene	103-65-1	-0.121	0.087	-0.261	1	0	
o-phenylenediamine	95-54-5	-0.452	0.100	0.200	2	0	1.37
o-toluenesulfonamide	88-19-7	-0.350	0.086	0.126	2	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC\text{-}1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Oxyfluorfen	42874-03-3	0.931	0.127	0.070	6	0	
p-cymene	99-87-6	0.433	0.069	-0.446	2	0	1.83
p-phenylenediamine	106-50-3	-0.526	0.113	0.254	2	0	**3.21**
Propoxur	114-26-1	0.678	0.098	-0.625	2	0	
Resorcinol	108-46-3	0.002	0.058	-0.131	2	0	
Sulphanilamide	63-74-1	-0.497	0.108	0.233	2	0	
Thiamphenicol	15318-45-3	-0.068	0.061	-0.080	2	0	
Toluene	108-88-3	-0.175	0.087	-0.222	1	0	0.57
Trans-cinnamic acid	140-10-3	-1.267	0.125	-0.051	1	1	
2-(1,1-dimethyl)-4,6-dimethylphenol	1879-09-0	1.411	0.128	-0.720	4	0	1.85
2,3,4,5-tetrachlorophenol	4901-51-3	0.391	0.112	0.244	5	0	
2,3,5,6-tetrachlorophenol	935-95-5	0.428	0.107	0.217	5	0	
2,3,5-trichlorophenol	933-78-8	0.264	0.070	0.117	4	0	
2,3,5-trimethylphenol	697-82-5	0.712	0.051	-0.210	4	0	
2,3,6-trichlorophenol	933-75-5	0.264	0.070	0.117	4	0	
2,3,6-trimethylphenol	2416-94-6	0.734	0.051	-0.226	4	0	
2,3-dimethylphenol	526-75-0	0.304	0.045	-0.132	3	0	
2,4,6-tribromophenol	118-79-6	0.344	0.063	0.059	4	0	2.34
2,4,6-trimethylphenol	527-60-6	0.863	0.056	-0.320	4	0	
2,4-dibromophenol	615-58-7	0.178	0.050	-0.040	3	0	1.84
2,4-di-*tert*-butylphenol	96-76-4	1.410	0.185	-0.939	3	0	2.48
2,5-dimethylphenol	95-87-4	0.516	0.046	-0.287	3	0	1.33
2,6-dimethylphenol	576-26-1	0.418	0.044	-0.215	3	0	0.91
2,6-di-*sec*-butylphenol	5510-99-6	1.079	0.108	-0.698	3	0	3.14
2-allylphenol	1745-81-9	0.099	0.056	-0.202	2	0	
2-ethylphenol	90-00-6	0.226	0.057	-0.295	2	0	
2-methoxyphenol	90-05-1	0.065	0.056	-0.177	2	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
2-*n*-propylphenol	644-35-9	0.265	0.059	-0.323	2	0	
2-*tert*-butyl phenol	88-18-6	0.499	0.075	-0.494	2	0	1.61
2-*tert*-butyl-*p*-cresol	2409-55-4	0.973	0.090	-0.620	3	0	1.94
3,4,5-trichloroguaiacol	57057-83-7	0.684	0.083	0.030	5	0	
3,4,5-trichlorophenol	609-19-8	0.223	0.075	0.147	4	0	
3,4-dimethylphenol	95-65-8	0.355	0.044	-0.169	3	0	
3-ethylphenol	620-17-7	0.254	0.058	-0.315	2	0	
3-trifluoromethyl-4-nitrophenol	88-30-2	0.211	0.048	-0.064	3	0	
4-(1-methylethenyl)phenol	4286-23-1	0.204	0.057	-0.279	2	0	1.16
4-(2,4-dichlorophenoxy)-phenol	40843-73-0	0.872	0.076	-0.107	5	0	
4,5,6-trichloroguaiacol	2668-24-8	0.597	0.089	0.094	5	0	
4,5-dichloroguaiacol	2460-49-3	0.361	0.062	0.046	4	0	
4-chloro-2-methylphenol	1570-64-5	0.164	0.050	-0.030	3	0	
4-chloro-2-nitrophenol	89-64-5	-0.111	0.077	0.171	3	0	1.20
4-ethylphenol	123-07-9	0.166	0.056	-0.251	2	0	
4-*n*-nonylphenol	104-40-5	0.570	0.083	-0.546	2	0	
4-*n*-octylphenol	1806-26-4	0.444	0.070	-0.454	2	0	3.37
4-pentylphenol	14938-35-3	0.263	0.059	-0.322	2	0	2.07
6-*tert*-butyl-*m*-cresol	88-60-8	0.955	0.087	-0.607	3	0	1.78
6-*tert*-butyl-*o*-cresol	2219-82-1	0.729	0.060	-0.442	3	0	1.58
Methyl *p*-hydroxybenzoate	99-76-3	0.147	0.056	-0.237	2	0	0.40
o-*sec*-butylphenol	89-72-5	0.348	0.063	-0.384	2	0	1.40
p-*sec*-butylphenol	99-71-8	0.191	0.057	-0.269	2	0	1.76
Tetrachloroguaiacol	2539-17-5	0.951	0.126	0.055	6	0	
Thymol	89-83-8	0.700	0.057	-0.421	3	0	1.50
Trichlorosyringol	2539-26-6	0.972	0.125	0.040	6	0	
1-nitropyrene	5522-43-0	-0.790	0.143	0.227	1	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
2-acetamidophenoxazin-3-one	1916-55-8	-0.129	0.064	-0.036	2	0	
2-amino-7-methoxyphenoxazin-3-one		0.408	0.044	-0.208	3	0	
9-vinylcarbazole	1484-13-5	-0.773	0.154	-0.005	0	0	
Acenaphthene	83-32-9	0.000	0.058	-0.130	2	0	
Dibenzo[b,f]cyclohepten-1-one	2222-33-5	0.474	0.054	-0.036	4	0	
Dibenzothiophene	132-65-0	-0.727	0.150	-0.039	0	0	**2.12**
Fluorene	86-73-7	0.457	0.055	-0.024	4	0	
Phenanthrene	85-01-8	-0.638	0.143	-0.104	0	0	**2.10**
Phenothiazine	92-84-2	0.523	0.052	-0.072	4	0	**2.41**
1-decanol	112-30-1	-0.097	0.149	-0.499	0	0	**1.75**
1-nonanol	143-08-8	-0.286	0.138	-0.361	0	0	**1.65**
2-(2-butoxyethoxy)ethanol	112-34-5	-0.290	0.138	-0.358	0	0	
2-butoxyethanol	111-76-2	-0.406	0.136	-0.273	0	0	
2-isoproxyethanol	109-59-1	-0.594	0.141	-0.136	0	0	
3-pentanol	71-41-0	-0.653	0.144	-0.093	0	0	
Cyclohexanol	108-93-0	-0.295	0.138	-0.354	0	0	
Hexanol	111-27-3	-0.566	0.140	-0.156	0	0	
Isodecyl alcohol	25339-17-7	-0.398	0.136	-0.279	0	0	**1.43**
1,6-hexanediamine	124-09-4	-0.572	0.140	-0.152	0	0	**0.21**
2-(dibutylamino)ethanol	102-81-8	-0.140	0.146	-0.467	0	0	**0.78**
2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)	6864-37-5	0.045	0.163	-0.602	0	0	**1.04**
2-amino-2-ethylpropanediol	115-70-8	-0.699	0.148	-0.059	0	0	
Cyclohexylamine	108-91-8	-0.350	0.137	-0.314	0	0	**0.48**
Dibutylamine	111-92-2	-0.129	0.146	-0.475	0	0	
Diethanolamine	111-42-2	-0.560	0.139	-0.161	0	0	
Diethylamine	109-89-7	-0.558	0.139	-0.162	0	0	**0.43**
Diethylnitrosamine	55-18-5	-0.369	0.136	-0.300	0	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Diisopropylamine	108-18-9	-0.075	0.151	-0.515	0	0	
Dimethylamine	124-40-3	-0.450	0.136	-0.241	0	0	
Piperidine	110-89-4	-0.249	0.140	-0.388	0	0	
t-butylamine	75-64-9	-0.412	0.136	-0.269	0	0	
Triethylamine	121-44-8	-0.362	0.136	-0.305	0	0	0.63
2-(dimethylamino)ethyl methacrylate	2867-47-2	-0.147	0.145	-0.462	0	0	0.92
2-ethylhexyl methacrylate	688-84-6	0.080	0.168	-0.628	0	0	1.85
2-hydroxyethyl acrylate	818-61-1	-0.599	0.141	-0.132	0	0	1.25
2-hydroxyethyl methacrylate	868-77-9	-0.480	0.137	-0.219	0	0	
Ethyl trichloroacetate	515-84-4	-0.683	0.147	-0.071	0	0	0.64
Ethylacrylate	140-88-5	-0.503	0.138	-0.202	0	0	1.92
Isobutyl acetate	110-19-0	-0.334	0.137	-0.326	0	0	0.83
Methyl acrylate	96-33-3	-0.490	0.137	-0.212	0	0	1.79
Methyl methacrylate	80-62-6	-0.387	0.136	-0.287	0	0	
n-butyl acrylate	141-32-2	-0.472	0.137	-0.225	0	0	1.73
n-butyl methacrylate	97-88-1	-0.251	0.139	-0.386	0	0	1.40
Testosterone propionate	57-85-2	0.475	0.241	-0.916	0	0	
Tetracaine	94-24-6	0.634	0.091	-0.593	2	0	
Vinyl acetate	108-05-4	-0.731	0.151	-0.036	0	0	1.55
2,3,4-trichloroaniline	634-67-3	0.029	0.102	0.289	4	0	
2,3-dichloroaniline	608-27-5	-0.170	0.086	0.214	3	0	
2,3-dimethylaniline	87-59-2	0.147	0.052	-0.017	3	0	
2,4,5-trichloroaniline	636-30-6	0.063	0.097	0.264	4	0	
2,4,6-trichloroaniline	634-93-5	0.115	0.089	0.226	4	0	1.57
2,4,6-trimethylaniline	88-05-1	0.633	0.050	-0.152	4	0	**0.39**
2,4-dichloroaniline	554-00-7	-0.145	0.082	0.196	3	0	1.30
2,4-dimethylaniline	95-68-1	0.173	0.050	-0.036	3	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (h*=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
2,5-dichloroaniline	95-82-9	-0.121	0.078	0.178	3	0	1.87
2,5-dimethylaniline	95-78-3	0.163	0.050	-0.029	3	0	
2,6-dichloroaniline	608-31-1	-0.122	0.079	0.179	3	0	
2,6-diethylaniline	579-66-8	0.284	0.045	-0.117	3	0	
2,6-dimethylaniline	87-62-7	0.162	0.051	-0.028	3	0	
2-chloro-5-methylaniline	95-81-8	0.097	0.055	0.019	3	0	1.11
2-chloroaniline	95-51-2	-0.350	0.086	0.126	2	0	1.24
2-ethylaniline	578-54-1	-0.161	0.067	-0.012	2	0	
2-methyl-4-nitroaniline	99-52-5	-0.116	0.078	0.175	3	0	
2-methylaniline	95-53-4	-0.286	0.078	0.079	2	0	-0.15
2-nitroaniline	88-74-4	-0.509	0.110	0.242	2	0	0.52
2-nitro-*p*-anisidine	96-96-8	-0.070	0.072	0.141	3	0	0.61
3,4,5-trichloroaniline	634-91-3	0.090	0.093	0.244	4	0	
3,4-dichloroaniline	95-76-1	-0.079	0.073	0.148	3	0	1.17
3,4-dimethylaniline	95-64-7	0.129	0.053	-0.004	3	0	
3,5-dichloroaniline	626-43-7	-0.103	0.076	0.165	3	0	
3,5-dimethylaniline	108-69-0	0.274	0.045	-0.110	3	0	0.55
3-chloroaniline	108-42-9	-0.313	0.082	0.099	2	0	1.16
3-ethylaniline	587-02-0	-0.035	0.059	-0.104	2	0	
3-methylaniline	108-44-1	-0.141	0.065	-0.027	2	0	
3-nitroaniline	99-09-2	-0.322	0.083	0.105	2	0	0.19
4-chloro-2-nitroaniline	89-63-4	-0.315	0.110	0.320	3	0	1.01
4-chloroaniline	106-47-8	-0.320	0.082	0.104	2	0	1.34
4-ethylaniline	589-16-2	-0.004	0.058	-0.127	2	0	
4-fluoroaniline	371-40-4	-0.309	0.081	0.096	2	0	
4-isopropylaniline	99-88-7	0.059	0.056	-0.173	2	0	0.47
4-methylaniline	106-49-0	-0.067	0.061	-0.081	2	0	-0.05

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
4-nitroaniline	100-01-6	-0.348	0.086	0.124	2	0	0.21
Aniline, *p*-(phenylazo)	60-09-3	-0.007	0.065	0.095	3	0	2.75
Chlorfluoroaniline	21397-08-0	-0.138	0.081	0.191	3	0	
Dithiodianiline	722-27-0	0.053	0.098	0.271	4	0	
n,*n*-Diethylaniline	91-66-7	0.205	0.103	-0.499	1	0	0.76
n,*n*'-dimethylaniline		-0.360	0.093	-0.087	1	0	
n,*n*-Dimethylaniline	121-69-7	-0.360	0.093	-0.087	1	0	
n-ethylaniline	103-69-5	-0.115	0.087	-0.266	1	0	0.22
p-anisidine	104-94-9	-0.083	0.062	-0.069	2	0	
2-ethyl butanoic acid	88-09-5	-1.405	0.175	-0.170	0	1	0.14
2-methyl butanoic acid	600-07-7	-1.518	0.179	-0.088	0	1	
3-methyl butanoic acid	503-74-2	-1.278	0.174	-0.263	0	1	
Acrylic acid	79-10-7	-1.629	0.186	-0.007	0	1	0.07
A-fluoro-b-alanine	3821-81-6	-1.698	0.192	0.044	0	1	
Chloroacetic acid	79-11-8	-1.678	0.190	0.029	0	1	0.12
Dansylglycine	1091-85-6	-0.885	0.085	-0.110	2	1	
Flumequine	42835-25-6	-0.011	0.078	-0.309	4	1	
Gentisic acid	490-79-9	-0.866	0.081	0.096	3	1	
Glyoxylic acid	298-12-4	-1.709	0.193	0.052	0	1	0.26
Heptanoic acid	111-14-8	-1.272	0.174	-0.267	0	1	0.24
Isocyanuric acid	108-80-5	0.303	0.045	-0.131	3	0	
Malonic acid diethylester	105-53-3	-0.720	0.150	-0.044	0	0	0.61
Methacrylic acid	79-41-4	-1.570	0.182	-0.050	0	1	
Octanoic acid	124-07-2	-1.140	0.179	-0.364	0	1	0.45
Orthoformic acid trimethylester	149-73-5	-0.387	0.136	-0.287	0	0	
Oxolinic acid	14698-29-4	-0.168	0.070	-0.194	4	1	
Perfluorooctanoic acid	335-67-1	-2.024	0.239	0.282	0	1	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Pivalic acid	75-98-9	-1.219	0.176	-0.306	0	1	
Pyridaphenthion	119-12-0	-0.198	0.087	-0.205	1	0	
Sorbic acid	110-44-1	-1.640	0.187	0.001	0	1	0.17
2-methylvaleraldehyde	123-15-9	-0.658	0.145	-0.089	0	0	
Capronaldehyde	66-25-1	-0.491	0.137	-0.211	0	0	
Crotonaldehyde	4170-30-3	-0.609	0.142	-0.125	0	0	**2.99**
Dodecanal	112-54-9	0.121	0.174	-0.658	0	0	
Glutaraldehyde	111-30-8	-0.725	0.150	-0.040	0	0	1.06
Propionaldehyde	123-38-6	-0.606	0.142	-0.127	0	0	
1,2,4-trichloro-5-nitrobenzene	89-69-0	-0.122	0.131	0.399	4	0	
2-chloro-1-fluoro-4-nitrobenzene	350-30-1	-0.001	0.064	0.091	3	0	1.94
2-chloronitrobenzene	88-73-3	-0.452	0.100	0.200	2	0	
2-nitroanisole	91-23-6	-0.160	0.067	-0.013	2	0	0.51
2-nitrotoluene	88-72-2	-0.383	0.090	0.150	2	0	
3,4-dichloronitrobenzene	99-54-7	-0.100	0.076	0.163	3	0	1.61
3-nitroanisole	555-03-3	0.059	0.056	-0.173	2	0	0.41
4-amino-2,6-dinitrotoluene	19406-51-0	-0.159	0.139	0.426	4	0	
4-amino-2-nitrotoluene	119-32-4	-0.305	0.108	0.313	3	0	
4-chloro-3-methylnitrobenzene	13290-74-9	0.088	0.056	0.026	3	0	
4-chloronitrobenzene	100-00-5	-0.261	0.076	0.061	2	0	
4-methylnitrobenzene	99-99-0	0.010	0.058	-0.137	2	0	0.57
alpha-chloro-4-nitrotoluene	100-14-1	-0.026	0.059	-0.111	2	0	2.45
Nitrobenzene	98-95-3	-0.415	0.097	-0.047	1	0	
1,3-diphenylguanidine	102-06-7	-0.055	0.060	-0.090	2	0	1.09
2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	901-44-0	1.565	0.163	-0.833	4	0	1.18
2,3,4,4'-tetrahydroxybenzophenon	31127-54-5	1.223	0.119	-0.143	6	0	0.84
2,4-diamino-6-phenyl-*s*-triazine	91-76-9	0.426	0.057	-0.001	4	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
2-hydroxy-4-methoxybenzophenone	131-57-7	0.883	0.058	-0.335	4	0	1.78
3,4,4'-trichlorodiphenylurea	101-20-2	0.641	0.085	0.062	5	0	
4,4'-diaminodiphenyl ether	101-80-4	0.566	0.051	-0.103	4	0	
4,4'-dihydroxydiphenylmethane	620-92-8	0.501	0.053	-0.056	4	0	1.19
4,4'-methylenedianiline	101-77-9	0.340	0.063	0.062	4	0	0.98
Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-	51963-82-7	1.183	0.088	-0.554	4	0	1.08
Benzophenone	119-61-9	-0.187	0.069	0.007	2	0	
Bis(4-hydroxyphenyl)sulfone	80-09-1	0.542	0.052	-0.086	4	0	
Bisphenol A	80-05-7	1.239	0.097	-0.595	4	0	1.46
Dibenzyl ether	103-50-4	0.400	0.066	-0.422	2	0	1.46
Dibromocresyl glycidyl ether	30171-80-3	0.121	0.056	-0.218	2	0	2.39
Diphenyl ether	101-84-8	0.306	0.060	-0.353	2	0	1.98
Diphenylamine	122-39-4	-0.045	0.060	-0.097	2	0	1.41
di-*p*-tolylamine	620-93-9	0.908	0.059	-0.353	4	0	2.66
Hydrazobenzene	122-66-7	-0.226	0.072	0.035	2	0	3.63
n,*n*'-bis(2-methylphenyl)guanidine	97-39-2	0.624	0.050	-0.146	4	0	1.10
Procymidone	32809-16-8	0.221	0.047	-0.071	3	0	
Pyrimethamine	58-14-0	1.227	0.119	-0.146	6	0	
Styrene-7,8-oxide	96-09-3	-0.449	0.099	-0.022	1	0	1.14
Sulfadiazine	68-35-9	0.056	0.058	0.049	3	0	
Sulfadimethoxine	122-11-2	0.809	0.077	-0.061	5	0	
Sulfamethazine	57-68-1	0.973	0.076	-0.181	5	0	
Sulfamethoxazole	723-46-6	0.357	0.062	0.049	4	0	
Tetrabromobisphenol A	79-94-7	2.023	0.276	-0.288	8	0	1.77
Trimethoprim	738-70-5	1.621	0.186	-0.214	7	0	
1,1'-oxybis-butane	142-96-1	-0.512	0.138	-0.196	0	0	
1,5-cyclooctadiene	111-78-4	-0.447	0.136	-0.243	0	0	0.92

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC\text{-}1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
1-benzo[b]thien-2-ylethan-1-one	22720-75-8	-0.528	0.107	0.036	1	0	1.25
1-chloro-2,4-dinitrobenzene	97-00-7	-0.253	0.099	0.275	3	0	**3.10**
1-cyclohexene-1-carbonitrile	1855-63-6	-0.544	0.139	-0.172	0	0	0.42
1-mercaptooctane [n-Octylmercaptan]	111-88-6	-0.227	0.140	-0.404	0	0	2.65
1-methoxy-2-propanol	107-98-2	-0.498	0.137	-0.206	0	0	
2-(1'-cyclohexenyl)cyclohexanone	1502-22-3	-0.168	0.144	-0.447	0	0	0.79
2,2,5,5,-tetramethylhydrofuran	15045-43-9	-0.551	0.139	-0.167	0	0	
2,2,6,6-tetramethylpiperidin-4-ol	2403-88-5	0.025	0.161	-0.588	0	0	
2,2'-dithiobisbenzothiazole	120-78-5	-0.323	0.083	0.106	2	0	
2,3,3,3,2',3',3',3'-octachlorodipropyl ether	127-90-2	-0.272	0.139	-0.371	0	0	2.81
2,3-dichloro-1,4-napthoquinone	117-80-6	-0.290	0.079	0.082	2	0	3.86
2,4,6-trimethylbenzaldehyde	487-68-3	0.782	0.053	-0.261	4	0	1.09
2-acetoxy-1,4-naphthoquinone	1785-65-5	-0.271	0.077	0.068	2	0	
2-aminophenol	95-55-6	-0.166	0.067	-0.009	2	0	2.21
2-aminopyridine	504-29-0	-0.364	0.093	-0.084	1	0	0.93
2-butanone oxime	96-29-7	-0.381	0.136	-0.291	0	0	
2-butenedinitrile, (E)-	764-42-1	-0.780	0.155	0.000	0	0	2.34
2-chlorobenzyl chloride	611-19-8	0.036	0.057	-0.156	2	0	2.78
2-decanone	693-54-9	-0.281	0.138	-0.364	0	0	
2-hydroxy-1,4-naphthoquinone	83-72-7	-0.327	0.083	0.109	2	0	
2-mercaptobenzothiazole	149-30-4	-0.248	0.074	0.051	2	0	
2-mercaptoethanol	60-24-2	-0.690	0.147	-0.066	0	0	0.43
2-methyl-1,4-naphthoquinone	58-27-5	-0.007	0.058	-0.125	2	0	
2-methyl-4-nitroimidazole	696-23-1	-0.071	0.061	-0.078	2	0	
2-methylthio-4-*tert*-butylamino-6-amino-*s*-triazine	30125-65-6	0.807	0.068	-0.499	3	0	
2-Oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	470-82-6	0.049	0.164	-0.605	0	0	
2-phenylindole	948-65-2	0.043	0.057	-0.161	2	0	2.85

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
2-phenylphenol	90-43-7	0.304	0.045	-0.132	3	0	
2-propenenitrile, 2-chloro-	920-37-6	-0.709	0.149	-0.052	0	0	3.02
2-propenenitrile	107-13-1	-0.686	0.147	-0.069	0	0	1.02
2-undecanone	112-12-9	-0.229	0.140	-0.402	0	0	
2-vinylpyridine	100-69-6	-0.146	0.087	-0.243	1	0	1.21
3-(methylthio)propionaldehyde	3268-49-3	-0.418	0.136	-0.264	0	0	1.17
3,3'-dichlorobenzidine	91-94-1	0.718	0.148	0.225	6	0	2.70
3,4,5-trichlorocatechol	56961-20-7	0.447	0.105	0.203	5	0	
3,4,6-trichlorocatechol	32139-72-3	0.406	0.110	0.233	5	0	
3,4-dichlorocatechol	3978-67-4	0.315	0.065	0.080	4	0	
3,5,5-trimethyl-2-cyclohexen-1-one	78-59-1	-0.101	0.149	-0.496	0	0	
3,5-dichlorocatechol	13673-92-2	0.271	0.070	0.112	4	0	
3,5-di-*tert*-butylsalicylic acid	19715-19-6	0.657	0.186	-0.796	4	1	1.97
3a,4,7,7a-tetrahydro-1H-indene	3048-65-5	-0.418	0.136	-0.264	0	0	1.44
3-amino-1,2,4-triazole	61-82-5	-0.472	0.101	-0.005	1	0	
3-amino-2-Butenenitrile	1118-61-2	-0.818	0.159	0.028	0	0	0.68
3-amino-4-chlorobenzoic acid	2840-28-0	-1.040	0.101	0.223	3	1	
3-aminophenol	591-27-5	-0.185	0.069	0.005	2	0	-0.04
3-aminopyridine	462-08-8	-0.369	0.094	-0.080	1	0	1.04
4,4,4-trifluorocrotonitrile	406-86-0	-0.614	0.142	-0.121	0	0	2.78
4,4'-dihydroxy-biphenyl	92-88-6	0.719	0.051	-0.215	4	0	1.16
4,5-dichlorocatechol	3428-24-8	0.334	0.064	0.066	4	0	
4,6-dinitro-*o*-cresol	534-52-1	0.268	0.070	0.114	4	0	2.26
4-amino-2-nitrophenol	119-34-6	-0.212	0.092	0.245	3	0	
4-aminophenol	123-30-8	-0.256	0.075	0.057	2	0	2.07
4-aminopyridine	504-24-5	-0.430	0.098	-0.036	1	0	1.44
4-chlorocatechol	2138-22-9	0.204	0.048	-0.059	3	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
4-ethyl-1,1'-biphenyl	5707-44-8	0.570	0.048	-0.326	3	0	2.48
4-hydroxybenzoic acid	99-96-7	-1.021	0.090	-0.011	2	1	0.17
4-methylbenzoic acid	99-94-5	-0.884	0.085	-0.111	2	1	0.33
4-vinylpyridine	100-43-6	-0.131	0.087	-0.254	1	0	**2.02**
5-ethylidene-8,9,10-trinorborn-2-ene	16219-75-3	0.180	0.182	-0.701	0	0	1.23
6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	91-53-2	0.841	0.072	-0.524	3	0	1.51
6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	2439-01-2	0.292	0.045	-0.123	3	0	
8-hydroxyquinoline	148-24-3	-0.127	0.087	-0.257	1	0	1.76
Acetophenone	98-86-2	-0.360	0.093	-0.087	1	0	
Acetylsalicylic acid	50-78-2	-0.992	0.089	-0.032	2	1	
alpha,alpha-dichlorotoluene	98-87-3	-0.335	0.092	-0.105	1	0	0.85
Anisol	100-66-3	-0.132	0.087	-0.253	1	0	
Benazolin ethyl	25059-80-7	0.225	0.047	-0.074	3	0	
Benzaldehyde	100-52-7	-0.334	0.092	-0.106	1	0	
Benzenethiol	108-98-5	-0.284	0.089	-0.142	1	0	**4.09**
Benzoyl-chloride	98-88-4	-0.434	0.098	-0.033	1	0	
Benzyl-chloroformiate	501-53-1	-0.354	0.093	-0.091	1	0	
Betariboacetate	13035-61-5	-0.549	0.139	-0.169	0	0	
Biphenyl	92-52-4	0.113	0.056	-0.212	2	0	1.60
Bis(2-chloroethyl) ether	111-44-4	-0.507	0.138	-0.199	0	0	
Bismerthiazol	79319-85-0	-0.536	0.139	-0.178	0	0	
Buparvaquone	88426-33-9	0.525	0.078	-0.513	2	0	
But-3-en-3-olide	674-82-8	-0.557	0.139	-0.163	0	0	0.98
Butachlor	23184-66-9	0.853	0.073	-0.533	3	0	3.05
Butylbenzyl phthalate	85-68-7	0.485	0.045	-0.264	3	0	
Captopril	62571-86-2	-1.094	0.182	-0.397	0	1	
Chlorohydroquinone	615-67-8	0.214	0.048	-0.066	3	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Cinmethylin	87818-31-3	0.910	0.139	-0.794	2	0	
Cotinine	486-56-6	0.013	0.090	-0.359	1	0	
Cyclohexanone oxime	100-64-1	-0.299	0.138	-0.351	0	0	
Cyclohexanone	108-94-1	-0.392	0.136	-0.283	0	0	
Cyclosulfamuron	136849-15-5	1.594	0.138	-0.634	5	0	
Diallyl phthalate	131-17-9	0.333	0.062	-0.373	2	0	**2.75**
Diethyl disulfide	110-81-6	-0.383	0.136	-0.290	0	0	1.01
Diethyl malonate	105-53-3	-0.709	0.149	-0.052	0	0	0.61
Diethyl phthalate	84-66-2	0.117	0.056	-0.215	2	0	
Diisobutyl phthalate	84-69-5	0.265	0.059	-0.323	2	0	1.92
Dimethyl disulphide	624-92-0	-0.458	0.137	-0.235	0	0	**1.93**
Dimethylformamide	68-12-2	-0.487	0.137	-0.214	0	0	
Dimethylnitrosamine	62-75-9	-0.657	0.145	-0.090	0	0	
Dinitramine	29091-05-2	0.355	0.062	0.051	4	0	
Diphenylpropanediol		0.111	0.056	-0.211	2	0	
Droperidol	548-73-2	0.192	0.079	0.170	4	0	
Enrofloxacin	93106-60-6	0.172	0.096	-0.442	4	1	
Ethanethiol	75-08-1	-0.514	0.138	-0.194	0	0	1.45
Glycidyl methacrylate	106-91-2	-0.603	0.142	-0.129	0	0	1.71
Haloxyfop R	72619-32-0	1.297	0.095	-0.417	5	0	
Hexamethylene diacrylate	13048-33-4	0.056	0.165	-0.610	0	0	2.77
Hydrogenatedbisphenol A	80-04-6	0.295	0.202	-0.785	0	0	1.10
Hydroquinone	123-31-9	0.048	0.057	-0.165	2	0	
Isoprene	78-79-5	-0.314	0.137	-0.340	0	0	0.66
Maleic anhydride	108-31-6	-0.688	0.147	-0.067	0	0	
Medazepam	2898-12-6	0.975	0.064	-0.402	4	0	
Methacrylonitrile	126-98-7	-0.569	0.140	-0.154	0	0	

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC-1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
Methyl isothiocyanate	556-61-6	-0.605	0.142	-0.128	0	0	**2.78**
Methylhydrazine	60-34-4	-0.716	0.149	-0.047	0	0	2.08
m-toluic acid	99-04-7	-0.926	0.086	-0.080	2	1	0.22
n-(*tert*-butyl)-2-benzothiazolylsulfenamide	95-31-8	0.024	0.090	-0.367	1	0	2.23
n,*n*-Dimethylhydrazine	57-14-7	-0.665	0.145	-0.084	0	0	
n-cyclohexyl-2-benzothiazolylsulfenamide	95-33-0	-0.272	0.089	-0.151	1	0	2.10
n-methyl-*n*,*n*-bis(2-dimethylaminoethyl)amine	3030-47-5	0.546	0.258	-0.968	0	0	
o-acetoacetotoluidide	93-68-5	-0.161	0.067	-0.012	2	0	
o-chlorobenzonitrile	873-32-5	-0.116	0.064	-0.045	2	0	0.57
Octanedinitrile	629-40-3	-0.277	0.138	-0.367	0	0	-0.61
Olaquindox	23696-28-8	0.178	0.056	-0.260	2	0	
o-tolidine	119-93-7	0.788	0.140	0.174	6	0	1.21
Pentane-1-thiol	110-66-7	-0.555	0.139	-0.164	0	0	**1.90**
Phenol,4,4',4''-ethylidynetris-	27955-94-8	1.383	0.125	-0.260	6	0	
Phthalic anhydride	85-44-9	-0.212	0.071	0.025	2	0	
Phthalonitrile	91-15-6	-0.107	0.063	-0.052	2	0	**0.75**
Pivaloyl chloride	3282-30-2	0.042	0.163	-0.600	0	0	
p-methoxybenzaldehyde	123-11-5	0.093	0.056	-0.198	2	0	**0.53**
p-phenylphenol	92-69-3	0.469	0.045	-0.252	3	0	**1.70**
Propyl gallate	121-79-9	0.609	0.050	-0.135	4	0	
Propyzamide	23950-58-5	0.403	0.044	-0.204	3	0	
Pyrazosulfuron ethyl	93697-74-6	1.758	0.171	-0.754	5	0	
Quinoline	91-22-5	0.206	0.057	-0.280	2	0	**0.30**
Secobarbital	76-73-3	-0.058	0.152	-0.527	0	0	
Simazine	122-34-9	0.501	0.046	-0.276	3	0	
Sulfaquinoxaline	59-40-5	0.334	0.044	-0.154	3	0	
tert-butyl 2-ethylperoxyhexanoate	3006-82-4	0.106	0.171	-0.647	0	0	1.67

Table E1. (Continued).

NAME	CAS	Pred. $pEC_{50,\ MTT[PLHC\text{-}1]}$ by Eq. 4.13 (mM)	Hat value (*h**=0.375)	Mor28e	NaasC	CATS2D_01_AN	pLC_{50} (mM)
tert-butylhydroperoxide	75-91-2	-0.494	0.137	-0.209	0	0	-0.02
Tetrachlorocatechol	1198-55-6	0.557	0.171	0.343	6	0	
Tetrachlorohydroquinone	87-87-6	0.679	0.153	0.254	6	0	
Tetrachlorophthalic anhydride	117-08-8	0.735	0.146	0.213	6	0	
Tetrahydromethylphthalic anhydride	11070-44-3	-0.472	0.137	-0.225	0	0	
Thiopental	76-75-5	-0.162	0.144	-0.451	0	0	
Thiophene	110-02-1	-0.665	0.145	-0.084	0	0	0.43
Thiosemicarbazide	79-19-6	-0.990	0.182	0.153	0	0	0.64
Thiourea dioxide	4189-44-0	-0.940	0.175	0.117	0	0	
Thiourea	62-56-6	-1.028	0.189	0.181	0	0	
Triclosan	3380-34-5	1.027	0.122	0.000	6	0	2.64
Trifluralin	1582-09-8	0.753	0.052	-0.240	4	0	
Trimethylquinone	935-92-2	-0.653	0.144	-0.093	0	0	
Triphenyl phosphate	115-86-6	0.823	0.069	-0.511	3	0	2.40
Tris-(2,3-dibromopropyl) phosphate	126-72-7	-0.369	0.136	-0.300	0	0	2.56
Tris(2-chloroethyl) phosphate	115-96-8	-0.525	0.138	-0.186	0	0	

*Bold values indicate pLC_{50} data with very weak relationship with the $pEC_{50,\ MTT[PLHC\text{-}1]}$ prediction.