ON THE AQUATIC TOXICITY OF DIVERSE CHEMICALS: DEVELOPMENT OF NOVEL *IN SILICO* MODELS TOWARDS SELECTED AQUATIC ORGANISMS UNDER THE FRAMEWORK OF REACH REGULATION

by

Serli Önlü

B.Sc. in Chemistry, Hacettepe University, 2005

M.Sc. in Chemistry, Boğaziçi University, 2008

Submitted to the Institute of Environmental Sciences in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in
Environmental Sciences

Boğaziçi University

2018

ACKNOWLEDGEMENTS

First and foremost, I am deeply indebted to Prof. Dr. Melek Türker Saçan, my dissertation advisor and mentor, for her invaluable support throughout my studies. Her wisdom always illuminated my way whenever I needed.

I wish to thank the members of my examining committee, Prof. Dr. Nadim Copty, Prof. Dr. Safiye Sağ Erdem, Prof. Dr. Meral Birbir, and Assoc. Prof. Dr. Egemen Aydın for advising and guiding my work.

I would like to thank Prof. Paola Gramatica for having provided the QSARINS software and Dr. Nikola Minovski for the fruitful scientific collaboration. I would like to take the opportunity to thank Mr. Vuranel Okay, on behalf of my former workplace Cosmetics and Cleaning Products Industry Association, for supporting me in pursuing my scientific ambitions.

I am sincerely grateful to my family for their immense love and support and to Dario for *the Pacific Ocean*.

It is not possible to acknowledge thoroughly all the people who had influenced me, to all those scientists, musicians, novelists, poets, and *to my beloved uncle*.

Last but not least, *to Alice*, the patient inner child, for being, and continuously educating me in every sense.

"For after all, as great scientists have said and as all children know, it is above all by the imagination that we achieve perception, and compassion, and hope."

Ursula K. Le Guin

This study was carried out in Ecotoxicology and Chemometrics Laboratory of Institute of Environmental Sciences, Boğaziçi University, İstanbul, Turkey. The financial supports of Boğaziçi University Scientific Research Funds (Project No: 8502 and 13463), the Scientific and Technological Research Council of Turkey (TÜBİTAK, Grant Number: 214Z225), and the Slovenian Research Agency (ARRS, Grant P1-0017) are appreciated.

ABSTRACT

ON THE AQUATIC TOXICITY OF DIVERSE CHEMICALS: DEVELOPMENT OF NOVEL *IN SILICO* MODELS TOWARDS SELECTED AQUATIC ORGANISMS UNDER THE FRAMEWORK OF REACH REGULATION

Environmental hazard and risk assessment of chemicals are crucial for aquatic species within the direction of the European Regulation on the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH). In consideration of ethical concerns, animal welfare and sustainability, as well as the need for aquatic toxicity data, *in silico* models, such as validated quantitative structure-toxicity relationships (QSTRs), are of great importance. In the present study, an activity-independent rational approach towards selecting an optimal geometry optimization method for improved QSTR modeling was proposed for the first time. Different QSTR and interspecies models towards three representative aquatic species (algae, fish, and planarian) were developed using the rational approach recommended. QSTR models on the prediction of cytotoxicity to rainbow trout liver cell line (RTL-W1), toxicity to *Dugesia japonica*, and an interspecies quantitative toxicity relationship between *Daphnia magna* and *D. japonica* were reported for the first time. The presented QSTR models have contributed to the literature by providing notable prediction coverage for environmentally significant chemicals, such as contaminants of emerging concern and high production volume chemicals. Furthermore, the first predicted aquatic toxicity and cytotoxicity data were provided for a great majority of the chemicals addressed in "The List of Chemicals with no Ecotoxicological Data" announced by the Scientific and Technological Research Council of Turkey (TÜBİTAK). The developed models are promising as potential tools in toxicity assessment, screening and prioritization of chemicals in a scientific and regulatory frame.

ÖZET

ÇEŞİTLİ KİMYASALLARIN SUCUL TOKSİSİTELERİ: REACH MEVZUATI ÇERÇEVESİNDE SEÇİLMİŞ SUCUL ORGANİZMALARA YÖNELİK YENİ MODELLER GELİŞTİRİLMESİ

Kimyasalların sucul organizmalara yönelik zararlılık ve risk değerlendirmesi Kimyasalların Kaydı, Değerlendirilmesi, İzni ve Kısıtlanması hakkındaki Avrupa Birliği Yönetmeliği (REACH) doğrultusunda önem taşımaktadır. Hayvan refahı ve sürdürülebilirlik gibi etik konular ile sucul toksisite veri gereksinimi göz önünde bulundurulduğunda, Kantitatif Yapı-Toksisitesi İlişkileri (KYTİ) gibi bilgisayarla modelleme yöntemleri çevresel toksisitelerin tahmin edilmesinde büyük önem taşımaktadır. Bu çalışmada, geometri optimizasyon yönteminin seçimine yönelik aktivite-bağımsız akılcı bir yaklaşım, iyileştirilmiş KYTİ modellemesi için ilk defa önerilmiştir. Temsili üç sucul türe (su yosunları, balık, planarya) yönelik farklı KYTİ ve türler arası modeller önerilen akılcı yaklaşım kullanılarak geliştirilmiştir. Gökkuşağı alabalığı karaciğer hücre dizini sitotoksisitesi ve *Dugesia japonica* toksisitesi için KYTİ modelleri ile *Daphnia magna* ve *D. japonica* için türler arası kantitatif toksisite ilişkisi ilk kez raporlanmıştır. Endişe verici kimyasallar ve yüksek hacimde üretilen kimyasallar gibi çevresel açıdan önemli kimyasallara yönelik geliştirilen ve dikkate değer tahmin kapsamı sağlayan KYTİ modelleri ile literatüre katkıda bulunulmuştur. Ayrıca, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından açıklanan "Ekotoksikolojik Verisi Olmayan Kimyasallar Listesi" nde belirtilen kimyasalların büyük bir çoğunluğu için tahmin edilen sucul toksisite ve sitotoksisite verileri ilk kez olarak sağlanmıştır. Geliştirilen modeller, bilimsel ve düzenleyici bir çerçevede kimyasalların toksisite değerlendirmesi, taranması ve önceliklendirilmesinde potansiyel araçlar olarak ümit vericidir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation	Unit
E	Gas-Phase Energy	eV
EC_{50}	Median Effective Concentration	
$EC_{50,\ AB}$	Decline of Cell Viability from AB Assay	μmol/L
$EC_{50,\ CFDA\text{-}AM}$	Decline of Cell Viability from CFDA-AM Assay	μmol/L
*ED**	Critical ED Value	
E_{HOMO}	The Highest Occupied Molecular Orbital Energy	eV
E_{LUMO}	The Lowest Unoccupied Molecular Orbital Energy	eV
*h**	Critical Hat Value	
F	Fischer Statistics	
K_{HSA}	HSA Binding Affinity Constant	
K_{ow}	n-Octanol/Water Partition Coefficient	
LC_{50}	Median Lethal Concentration	
Log K_{ow}	Logarithm of K_{ow}	
Log K_{HSA}	Logarithm of K_{HSA}	
Q^2_{LOO}	Leave-One-Out Cross Validation	
Q^2_{LMO}	Leave-Many-Out Cross Validation	
r	Pearson Correlation Coefficient	
R^2	Coefficient of Determination	
R^2_{adj}	Adjusted (for degrees of freedom) R^2	
48-h EC_{50}	48-hours *D. magna* Acute Immobilization	μmol/L
72-h EC_{50}	72-hours Algal Growth Inhibition	mol/L
48-h LC_{50}	48-hours *D. japonica* Mortality	μmol/L
96-h LC_{50}	96-hours Fish Acute Toxicity	μmol/L

Abbreviation	**Explanation**
AB	Alamar Blue
AD	Applicability Domain
ANN	Artificial Neural Networks
AS	All Subsets
CAS	Chemical Abstracts Service
CCC	Concordance Correlation Coefficient
CEC	Contaminants of Emerging Concern
CFDA-AM	5-Carboxyfluorescein Diacetate Acetoxymethyl Ester
CPANN	Counter Propagation Artificial Neural Network
DFT	Density Functional Theory
DS1	Dataset 1
DS2	Dataset 2
DS3	Dataset 3
EC	European Commission
ECHA	European Chemicals Agency
ED	Euclidean Distance
EDC	Endocrine Disrupting Chemical
EINECS	European Inventory of Existing Commercial Chemical Substances
EU	European Union
GA	Genetic Algorithm
GETAWAY	Geometry, Topology, and Atom-Weights Assembly
GM	Global Model
HF	Hartree-Fock
HPV	High Production Volume
HSA	Human Serum Albumin
ICE	Interspecies Correlation Estimation
KKDİK	Kimyasalların Kaydı, Değerlendirilmesi, İzni ve Kısıtlanması Hakkında Yönetmelik
LM1	Local Model 1
LM2	Local Model 2
MAE	Mean Absolute Error
MCDM	Multi-Criteria Decision Making
MLR	Multiple Linear Regression
MoA	Mode of Action

Abbreviation	**Explanation**
MoEU	TC Ministry of Environment and Urbanization
MSE	Mean Squared Error
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PC1	First Axis Principal Component
PM6	Parameterized Model 6
PPCPs	Pharmaceuticals and Personal Care Products
PRESS	Predicted Residual Sum of Squares
QSAR	Quantitative Structure-Activity Relationship
QSTR	Quantitative Structure-Toxicity Relationship
QTTR	Quantitative Toxicity-Toxicity Relationship
QUIK	Q^2 Under Influence of *K*
RDF	Radial Distribution Function
REACH	European Regulation for the Registration, Evaluation, Authorization and Restriction of Chemicals No 1907/2006
RSS	Residual Sum of Squares
RTL-W1	Rainbow Trout Liver Cell Line
SIDS	Screening Information Data Set
SM	Supplementary Material
SSE	Sum of the Squared Errors
TC	Türkiye Cumhuriyeti
TG	Test Guideline
TPA	Tonne per Annum
TSR	Training Set Range
TSS	Total Sum of Squares
TÜBİTAK	Scientific and Technological Research Council of Turkey
US EPA	United States Environmetal Protection Agency
WHIM	Weighted Holistic Invariant Molecular
3D-MoRSE	3D-Molecule Representation of Structures based on Electron Diffraction

1. INTRODUCTION

The chemical universe is immense and expanding. The advent of modern industrialization and technology allow human ingenuity to produce an enormous amount of chemicals: the Chemical Abstracts Service (CAS) Registry has over 141 million entries at the present. Strikingly, almost 388,000 chemicals are regulated or inventoried to date, such as the European Regulation for the "Registration, Evaluation, Authorization and Restriction of Chemicals No 1907/2006 (REACH)" List of Registered Substances, European Inventory of Existing Commercial Chemical Substances (EINECS) and high production volume (HPV) Chemicals (CAS, 2018). The continuing emergence of new chemicals as well as the tremendous gap between existing and controlled chemicals point to a large number of chemicals to be potentially regulated in the future. HPV chemicals are substances with high regulatory concern and annual production or import volumes exceeding 1000 tonnes in at least one member country of the Organization for Economic Co-operation and Development (OECD) (OECD, 2009). Interestingly, despite their abundance, basic toxicity assessment data are still needed for many of them (Judson et al., 2009). Contaminants of emerging concern (CEC), such as pharmaceuticals and personal care products (PPCPs), pesticides, and endocrine disrupting chemicals (EDCs) are chemical substances that have been recently recognized for their environmental significance. Importantly, environmental impacts of the CEC have not been adequately explored and information on spatial and toxicological assessment is required, particularly for key geographical regions and representative species (Petrie et al., 2015).

Information on the aquatic toxicity of chemicals is of particular interest in the regulatory context due to the vital importance of the water compartment and its continuous exposure to a large number of chemicals. REACH aims at ensuring a high level of protection of the environment from the risks posed by the chemicals with annual production or import volumes of at least one tonne within the European Union (EU) (EC, 2006). Consequently, the industry is obliged to provide information on the toxicity assessments towards selected aquatic species, such as algae, invertebrates, and fish, to the European Chemicals Agency (ECHA). Priority is given to HPV chemicals under REACH. Moreover, national bylaws adapted from REACH, such as the Turkish "Kimyasalların Kaydı, Değerlendirilmesi, İzni ve Kısıtlanması Hakkında Yönetmelik (KKDİK)" (TC, 2017), are being enacted worldwide, suggesting an increasing global need for aquatic toxicity assessments towards representative trophic levels.

Algae, aquatic invertebrates (preferably *Daphnia*), and fish are recognized organisms representing different trophic levels in classical toxicity evaluations. Algae are primary producers providing nutrients and oxygen for higher trophic levels, thus, fundamental for the sustainability of the ecosystem (Boyce et al., 2010). Aquatic invertebrates, such as *Daphnia*, are crucial for food-web interactions; both as primary consumers and food sources for secondary consumers, transferring energy from lower to higher trophic levels (Stollewerk, 2010). *Dugesia japonica* is another aquatic invertebrate widely distributed in freshwater environments in East Asia acting as omnivore and detritivore, thus, has a critical ecological importance (Kawakatsu et al., 1995). Despite being a simple organism, it has biochemical and physiological properties comparable to higher vertebrates: a basal evolutionary position with a well-organized central nervous system and a brain homologous to mammalian (Buttarelli et al., 2008). Recently, it has emerged as an ideal alternative to *Daphnia* due to its unique properties for environmental toxicology studies (Li, 2008; Li, 2012a; Li, 2012b; Li, 2013a; Li, 2013b). Likewise, fish are critical species for ecosystem function and biodiversity as they are both prey and predator (Helfman et al., 2009).

Conventional toxicity tests have been dependent on intact animals that are exposed to different concentrations of chemicals. The amount of a chemical having a definite effect on 50% of the organisms after a specified test duration is the measure of acute toxicity endpoint. These endpoints are referred to as median lethal concentration that kills 50% of the organisms (LC_{50}) or median effective concentration that results in a certain effect, such as growth inhibition or immobilization, on 50% of the organisms (EC_{50}). OECD published specific test guidelines (TGs) covering three representative trophic levels to standardize toxicity testing. OECD TG 201, 202 and 203 set out the *in vivo* methodologies to measure the toxicity as 72-hours algal growth inhibition (72-h EC_{50}), 48-hours *Daphnia* acute immobilization (48-h EC_{50}), and 96-hours fish acute toxicity (96-h LC_{50}), respectively (OECD, 2006; OECD, 2004; OECD, 1992). Besides, *in vitro* assays using rainbow trout liver cell line (RTL-W1) (Brinkmann et al., 2014) for cytotoxicity assessment, such as Alamar Blue (AB) for changes in energy metabolism and 5-carboxyfluorescein diacetate acetoxymethyl ester (CFDA-AM) for evaluating membrane integrity (Schreer et al., 2005), are advantageous for minimizing animal use. They allow a cost-effective option showing high positive correlations with *in vivo* results and could be suited for the first screening of acute toxicity of environmentally significant chemicals on non-target organisms (Castaño et al., 2003). AB and CFDA-AM assays measure the effective concentration causing 50% decline of cell viability ($EC_{50,\ AB}$ and $EC_{50,\ CFDA\text{-}AM}$).

However, ethical issues arising in the last quarter century, such as animal welfare and sustainability, urge modern toxicology to resolve the toxicity assessments based on alternative approaches. *In silico* approaches, such as quantitative structure-activity/toxicity relationship (QSAR/QSTR) models, relating chemical structures to biological activity by means of mathematical and statistical modeling are of great importance for predicting environmental toxicities. The quality of QSTR models depends on their constitutive elements, such as the quality of the activity data and the applied statistical procedure as well as accurate molecular descriptors. QSTRs developed and validated in accordance with the OECD principles (OECD, 2007) are well recognized under REACH for providing reliable toxicity data without the need for *in vivo* and *in vitro* experiments (EC, 2006). Likewise, quantitative toxicity-toxicity relationship (QTTR) models generated with interspecies toxicity data are promising "green" alternatives to the evaluation of chemicals missing experimental data. Based on extrapolation, from the results of "tested" to "untested", such models can allow producing information using the existing data on surrogate species, such as *Daphnia magna* (Cronin, 2010).

Consequently, *in silico* modeling the toxicity of diverse chemicals towards environmentally significant aquatic species would provide invaluable information, reducing the demands for *in vivo* and *in vitro* testing, as well as providing data for toxicity assessment, screening and prioritization of chemicals in a scientific and regulatory frame.

1.1. The Objectives and Contributions of the Thesis

The objectives of this study are multi-fold. Firstly, to systematically investigate the effect of different quantum chemical methods used for geometry optimization on theoretical molecular descriptors and statistical quality of QSAR/QSTR models using structurally heterogeneous chemicals. The second objective is to provide a general rational approach intended for selecting an optimal geometry optimization method for improved modeling. Thirdly, to present comprehensive compilations of the available experimental and predicted ecotoxicity data on environmentally important chemicals towards selected aquatic organisms. The fourth objective is to develop validated QSTRs and an interspecies model for the prediction of acute toxicity and cytotoxicity towards three representative aquatic species (algae, fish, and planarian) using the knowledge gained from the analysis carried out in accordance with the first two objectives. Finally, to report *in silico*-predicted toxicity values for numerous environmentally significant chemicals with no toxicity data utilizing the developed QSTR models.

The contributions of this thesis are as follows:

- A general rational approach towards selecting an optimal geometry optimization method for improved QSAR/QSTR modeling was proposed.
- Comprehensive compilations of the experimental and predicted ecotoxicity data of environmentally significant chemicals for the representative aquatic organisms were presented.
- Validated multiple linear regression (MLR) models having wide applicability domains were developed for acute toxicity prediction towards mixed algae species (predominantly *Pseudokirchneriella subcapitata*) and *D. japonica*, using the rational approach recommended.
- A validated non-linear (counter propagation artificial neural network (CPANN)) model was developed for acute toxicity prediction of mixed algae species (predominantly *P. subcapitata*).
- A validated MLR model with wide applicability domain was developed for cytotoxicity prediction of RTL-W1 cell line, using the rational approach recommended.
- Possible use of cytotoxicity values for the estimation of acute fish toxicity was evaluated.
- Interspecies quantitative toxicity relationship between *D. magna* and *D. japonica* was investigated and a validated QTTR model was presented.
- Relevant ecotoxicological data gaps for environmentally significant chemicals were determined and analyzed.
- Predicted toxicity data to mixed algae species and *D. japonica*, and predicted cytotoxicity data to RTL-W1 were reported for a wide range of industrial chemicals (including HPV chemicals) with no data, using the developed models.
- Predicted toxicity and cytotoxicity data were reported for a great majority of the chemicals addressed in "The List of Chemicals with no Ecotoxicological Data" (SU0303, 2015) announced by the Scientific and Technological Research Council of Turkey (TÜBİTAK).
- In conclusion, validated alternative methods for ecotoxicity assessment and screening and prioritization of chemicals in a scientific and regulatory frame were presented. The data gaps in these fields were filled substantially.

2. THEORETICAL BACKGROUND

2.1. Environmentally Significant Chemicals

2.1.1. High Production Volume Chemicals

OECD is an intergovernmental organization that harmonizes policies to address international problems. In 1987, a systematic study of the safety assessment of existing chemicals was initiated by OECD (OECD, 1987). In 1991, OECD HPV Chemicals Program was launched to further these efforts, where the production volume was used as a surrogate for data on occupational, consumer, and environmental exposure (OECD, 1991). HPV chemicals are defined as chemicals produced/imported at levels greater than 1,000 tonnes per year in at least one OECD member country (OECD, 2009). From an environmental point of view, HPV chemicals are substances with large enough industrial use and release to potentially pose an environmental risk. To mitigate this risk, each member country agreed to sponsor a proportion of the HPV chemicals to be assessed based on a minimum set of information requirements (the Screening Information Data Set (SIDS)). In this context, ecotoxicity and environmental fate assessments were covered under the SIDS consisting of a relatively restricted set of data elements. Despite the progress made towards the environmental safety evaluation of HPV chemicals, noticeably, basic toxicity data are still a need for a significant fraction of HPV chemicals (Judson et al., 2009). However, the actions taken under the scope of the OECD HPV Chemicals Program provided a basis for the prioritization of chemicals with high regulatory concern. The most updated list of HPV Chemicals of OECD contains 4,638 chemicals (OECD, 2009).

2.1.2. Contaminants of Emerging Concern

CEC are chemical substances that have been recently recognized for their environmental significance. They can be newly appearing chemicals whose presence were not known before, existing chemicals with previously unexplored environmental effects, or long-known substances like HPV chemicals with recently discovered information on their environmental risks (Sauvé and Desrosiers, 2014). CEC cover several types of widely used industrial chemicals, such as PPCPs including synthetic musks, EDCs including synthetic and natural hormones, insecticides, pesticides, and surfactants. In recent years, concerns over the environmental impacts of CEC have been raising among the scientific and regulatory authorities due to their ubiquitous, yet, low detectable levels,

potential toxicity, and persistent, bioaccumulative and biomagnifying properties (Petrie et al., 2015). Given the fact that even the long-known HPV chemicals can lack basic toxicity data, identification, safety assessment, and prioritization of the CEC remain vague. Importantly, the environmental impacts of the CEC have not been adequately explored and information on spatial and toxicological assessment is required, particularly for key geographical regions and representative species (Petrie et al., 2015; Sauvé and Desrosiers, 2014).

2.2. REACH Regulation

In December 2006, the European Commission (EC) introduced a comprehensive legislation to address the use and impact of chemicals on the human health and the environment (EC, 2006). REACH was designed to ensure a high level of protection of the environment from the risks that can be posed by chemicals and to promote the use of alternative methods. HPV chemicals were given priority by covering the scope of the OECD HPV Chemicals Program under the REACH framework. According to the legislative measures, in order to place and keep substances on the EU market, the industry is obliged to register their chemicals with the ECHA by submitting information on the intrinsic properties, following clearly defined information requirements that are tonnage, hazard and risk related. Consequently, REACH requires industry to provide ecotoxicological information (acute toxicity to algae, fish and aquatic invertebrates, chronic toxicity to fish and aquatic invertebrates, sediment toxicity, acute and chronic toxicity to terrestrial invertebrates, plants and vertebrates) for the chemicals with annual production volumes of at least one tonne. To date around 21,000 unique substances have been registered with the ECHA (ECHA, 2018).

2.2.1. Information Requirements for Aquatic Toxicity

Aquatic toxicology is the study of the effects of manufactured chemicals and other anthropogenic and natural materials and activities on aquatic organisms at various levels. Aquatic toxicology is a multidisciplinary field integrating aquatic toxicology, ecology, and chemistry.

Information on the toxicity of chemicals for aquatic organisms has been given a particular importance in the regulatory context, as the aquatic compartment is a typical sink for industrial pollution due to direct releases and indirect emission pathways. Toxicity information is used for assessing the hazards and risks of substances to aquatic organisms. Moreover, data derived from toxicity on aquatic species can also be used as a basis for extrapolation to other compartments, such as sediment and soil.

For environmental safety evaluation of industrial chemicals, toxicity results of the following groups of aquatic organisms from different food-web levels are recommended: algae, invertebrates, and fish. The choice of these three trophic levels as primary producers and primary and secondary consumers is considered indicative in protecting the aquatic ecosystems. Type and extent of aquatic toxicity information required within the scope of REACH are determined by the annual production volume. The minimum aquatic toxicity information requirements, defined by the *"standard information requirements"*, are presented in Table 2.1 (EC, 2006; Sobanska et al., 2014; Tarazona et al., 2014).

Table 2.1. Standard information requirements for aquatic toxicity.

Information requirement	**Type[a]**	**1-10 tpa[b]**	**10-100 tpa**	**100-1,000 tpa**	**>1,000 tpa**
Short-term toxicity testing on invertebrates (preferred species *Daphnia*)	M	Y	Y	Y	Y
Growth inhibition study on aquatic plants (algae preferred)	M	Y	Y	Y	Y
Short-term toxicity testing on fish	M		Y	Y	Y
Long-term toxicity testing on invertebrates (preferred species *Daphnia*)	TP			Y	Y
Long-term toxicity testing on fish	TP			Y	Y

[a]M: Mandatory. TP: The data should be included if available. A testing proposal should be submitted to ECHA before conducting new studies. Y: Yes. [b]Tonne per annum.

2.2.2. Aquatic Toxicity Assessments

Toxicity tests have traditionally been conducted on the intact species where the representative species are exposed to different concentrations of chemicals. The concentration of a chemical having a certain effect, such as growth inhibition or immobilization, on the 50% of the organisms after a definite test duration is a measure of acute toxicity endpoint by means of LC_{50} or EC_{50}, respectively.

Competent authorities request valid and reliable toxicity data to regulate the use and impact of chemicals on the environment. Conventionally, these data have been produced by testing chemicals in accordance with a number of well-defined standardized protocols. To this end, OECD has developed and published specific TGs for three representative aquatic species: OECD TG 201, 202 and 203 specify the *in vivo* methodologies to measure the toxicity as 72-h EC_{50} algal growth inhibition, 48-h EC_{50} *Daphnia* acute immobilization, and 96-h LC_{50} fish acute toxicity, respectively (OECD, 2006; OECD, 2004; OECD, 1992). The recommended species for each TG cover the following: *P. subcapitata* and *Desmodesmus subspicatus*, *D. magna*, *and Oncorhynchus mykiss* (rainbow trout). In addition to conventional tests, *in vitro* assays using RTL-W1 for cytotoxicity assessment, such as AB measuring the changes in energy metabolism and CFDA-AM evaluating

membrane integrity, have been used (Brinkmann et al., 2014; Schreer et al., 2005). They allow minimizing animal use and a cost-effective alternative that has high positive correlations with *in vivo* results (Castaño et al., 2003).

Nevertheless, due to the long-standing public and governmental pressures arising from the 3Rs principles (Reduce, Replace, and Refine) proposed by Russel and Burch (1959), modern toxicology has been in search of alternative approaches for the replacement of animal tests. Accordingly, to avoid conducting tests particularly on vertebrate animals, REACH provides the possibility to use alternative methods to fill in the data requirements. In this context, QSARs and QSTRs developed and validated in accordance with the OECD principles (OECD, 2007) are well recognized under REACH for providing reliable toxicity data without the need for *in vivo* and *in vitro* experiments (EC, 2006).

2.2.3. Use of QSARs under the Framework of REACH Regulation

Innovation and the use of alternative methods are one of the essential components highlighted in REACH. Among the alternative methods, QSARs (QSTRs in case of toxicity assessments) are mentioned within REACH under several points. Annex XI provides a broad legislative framework for fulfilling the information requirements while limiting vertebrate testing to the extent possible. Article 25 states that in order to avoid animal testing; testing on vertebrate animals shall be undertaken only as a last resort. This includes the need to gather all existing information on physicochemical, toxicological and ecotoxicological properties of a substance, including information generated by QSARs. In particular, Article 13 states that information on intrinsic properties of substances may be generated by means other than tests, through the use of alternative methods, for example qualitative or quantitative structure-activity relationship models. In Annex XI, the use of QSAR methods is foreseen when *"testing does not appear scientifically necessary"*, since the same level of information can be produced by means other than (vertebrate) testing. The following statements are written in the legal text regarding the use of QSARs:

"Results obtained from valid qualitative or quantitative structure-activity relationship models (QSARs) may indicate the presence or absence of a certain dangerous property. Results of QSARs may be used instead of testing when the following conditions are met:

- Results are derived from a QSAR model whose scientific validity has been established,
- The substance falls within the applicability domain of the QSAR model,

- Results are adequate for the purpose of classification and labeling and/or risk assessment,
- Adequate and reliable documentation of the applied method is provided."

An analysis of the ecotoxicity data submitted within the framework of REACH was performed by Sobanska and co-workers (2014). The results indicate that, from 1 June 2008 until 28 February 2011 (three months after the first registration deadline), 24,560 registration dossiers for 4,599 substances were successfully submitted to ECHA. Among them, 2,887 substances were evaluated to be "regular" registration dossiers and contain information on ecotoxicology. For 2-3% of these substances, ecotoxicity values were derived based on QSAR results, (2.6% for fish, 1.8% for aquatic invertebrates, and 2.1% for algae and aquatic plants), indicating that QSAR-generated data can be surely used for the information requirements under REACH Regulation and there is a prospective need for data generation by this means.

2.2.4. Regulatory Acceptance: The OECD Principles of QSAR Validation

The scientific validity of QSAR models should be proven for regulatory acceptance. Following the considerable international efforts towards the acceptability of QSAR models for regulatory purposes, OECD developed an internationally recognized set of principles for QSAR validation (OECD, 2007). Thus, the OECD principles of QSAR validation define the required criteria for the scientific validity of QSAR models:

"To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with the following information:

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."

Principle 1 states that the modeled endpoint should be well defined to ensure transparency in the endpoint that will be predicted by the model, since a certain endpoint could be determined by different protocols or experimental conditions. Principle 2 highlights that the reproducibility of the results is required; hence, the QSAR algorithm should be unambiguous. According to Principle 3, in order to provide information on the extent of predictive capability of the model, it is demanded to

define a domain of applicability based on the training set. Principle 4 addresses that a QSAR model must have been subjected to appropriate statistical validation procedures in order to determine the internal and external performance of the model. While internal validation represents goodness-of-fit and robustness and determined by the training set, external validation relates the predictivity of the model and determined by the test set. Finally, Principle 5 encourages the mechanistic interpretation of the model descriptors in relation to the modeled activity.

2.2.5. KKDİK Regulation: An Adaptation of REACH

To enable a smooth transition from the existing chemicals legislation to REACH, the EC has developed a number of REACH Implementation Projects to ensure that all stakeholders, particularly the industry and the public authorities, are adequately prepared for the practical application of the new system. In this direction, following the completion of the project entitled "Technical Assistance for Implementation of REACH Regulation in Turkey", Turkey has adapted a national bylaw. The Turkish KKDİK was published in June 2017 to improve the protection of the human health and the environment (TC, 2017). The executive and legislative authority for KKDİK is the TC Ministry of Environment and Urbanization (MoEU). In parallel with REACH, KKDİK requires the industry to register their chemicals with the MoEU, fulfilling the same set of information requirements defined by REACH. Consequently, information on aquatic toxicity towards algae, fish, and aquatic invertebrates for the chemicals with annual production or import volumes of at least one tonne should be provided, suggesting the need for aquatic toxicity data. Likewise, validated QSTRs are accepted for generating toxicity data for regulatory purposes.

2.3. Quantitative Structure Activity/Toxicity Relationships

Quantitative structure-activity relationships (QSARs) are *in silico* tools that relate chemical structures to biological activity by means of mathematical and statistical modeling. QSARs are advantageous because they rationalize a significant number of experimental observations, reduce animal testing and save resources. QSARs have been developed and used for over a half-century, starting with the pioneering works of Hansch and Fujita (1964) and Free and Wilson (1964). While Hansch analysis utilized hydrophobic, electronic and steric parameters as physicochemical descriptors, Free-Wilson analysis is based on structural fragments. Both studies provide a basis of contemporary QSAR research. Since then, a large number of QSARs developed for the prediction of a plenty of biological endpoints, using various statistical techniques (Cherkasov et al., 2014).

QSAR refers to quantitative structure-toxicity relationship (QSTR) when the biological activity is toxicity. The fundamental principle of QSTR is that toxicity is determined by chemical structure. Therefore, a basic QSTR has the following mathematical description, where toxicity is a function of structural properties:

$$\text{Toxicity} = f(\text{Structural properties}) + \text{Constant} \quad (2.1)$$

Structural properties are intrinsic properties of a chemical substance and are expressed via molecular descriptors. The constitutive elements of a QSTR model are experimental toxicity data, molecular descriptors and a mathematical/statistical function or algorithm that relates toxicity to structural properties. QSTR model development consist of the following steps: selection and preparation of the toxicity dataset (dependent variables), calculation of the molecular descriptors (independent variables), training set/test set division, selection of the descriptors best explaining the toxicity, construction of the model, model validation, and definition of the applicability domain (AD) of the model.

2.3.1. Toxicity Dataset

The appropriate selection of experimental toxicity dataset is a very important step for successful QSTR modeling. Moreover, the first principle of OECD validation criteria states the necessity of *"a defined endpoint"*. A modelable dataset should have certain characteristics, such as a proper diversity in chemical structures, a normal distribution of the toxicity data and the absence of activity cliffs (Golbraikh et al., 2014). Chemical identities in a proper toxicity dataset should be structurally diverse and free of salts, inorganic and organometallic compounds and mixtures (Tropsha, 2010). The fundamental principle behind QSTR modeling is that similar compounds possess similar toxicity. Introduced by Maggiora (2006), *"activity cliffs"* mean the condition that similar molecules may have very different toxicity values, implying that small changes in descriptor values may result in large changes in toxicity. Thus, in the case of a continuous response variable, large gaps exceeding 10% to 15% of the entire toxicity range are not recommended between two consecutive toxicity values ranked in ascending or descending order to avoid such *"toxicity cliffs"* (Tropsha, 2010). Consequently, the experimental toxicity dataset should be curated prior to model development to meet the above conditions.

2.3.2. Molecular Descriptors

Molecular descriptors are quantitative parameters defining the intrinsic molecular properties of a chemical structure. A highly recognized definition of a molecular descriptor was given by Todeschini and Consonni (2009): "The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment". This definition states that molecular descriptors are formal mathematical representations of a molecule, depicting chemical information encoded within the molecular structure. Molecular descriptors can be empirical obtained via experiments, or theoretical, calculated from a representation of the molecule, carrying different types of information obtained using algorithms. Only theoretical molecular descriptors were used in this study, thus, hereinafter, theoretical molecular descriptors will be referred to as molecular descriptors or descriptors for simplicity. Descriptors are usually classified as quantum chemical, physicochemical, structural, topological, electronic, and geometric. Another classification can be made based on different representations of the chemicals structure (Table 2.2).

Table 2.2. Classification of molecular descriptors based on structural representation.

Descriptor class	Representation	Information embedded
Zero-dimensional (0D)	Molecular formula	Constitutional descriptions and atomic properties. No information regarding bonds and bonding is expressed (0D).
One-dimensional (1D)	List of structural fragments, such as functional groups, fragments, and substituents	Presence/absence or count of different substructures, functional groups, or substituents. Connectivity between atoms (bonding) is expressed (1D).
Two-dimensional (2D)	Molecular graph or topological	Topological and connectivity indices, pairs of atoms at a certain topological distance. Atomic properties and connectivity between atoms (bonding) are expressed (2D).
Three-dimensional (3D)	3D geometrical	Geometrical properties, relative positions of atoms, substructures, functional groups, or substituents in 3D space. Atomic properties and connectivity between atoms (bonding) in 3D space are expressed.

2.3.3. Calculation of Molecular Descriptors

The quality of QSTR models depends on their constitutive elements, such as the quality of toxicity data and applied statistical procedure. However, it essentially begins with accurate molecular representations. 2D and 3D representations of molecular structures are simulated by means of dedicated software, such as SPARTAN (SPARTAN v. 10, 2011). Molecular descriptors are then calculated based on molecular representations via specific software. DRAGON is the most used application for the calculation of molecular descriptors (DRAGON v. 6, 2013). However,

molecules can exist and interact in a variety of conformations. Selection of the appropriate conformer(s) in QSTR studies is as critically important as the selection of appropriate molecular parameters, because different conformers obtained from different representations may significantly affect the estimation of their molecular descriptors. Therefore, obtaining accurate structural representations is a fundamental prerequisite for the development of reliable QSAR models, especially for those built with quantum chemical and 3D molecular descriptors (Önlü and Saçan, 2017a).

Quantum chemical computations are capable of generating intrinsic molecular properties that define the geometry of a chemical. They also provide thermodynamic and physicochemical data obtained from the electronic structure of molecules (Schüürmann, 2004; Enoch, 2010). Geometry optimization is a quantum chemical procedure locating the minimum energy, i.e., the most stable representation of a chemical structure (Schlegel, 2011). There are various quantum chemical calculation theories using different mathematical approximations for geometry optimization and derivation of quantum chemical descriptors. Semi-empirical methods, such as Parameterized Model 6 (PM6) (Stewart, 2007), deal only with valance electrons and use adjustable parameters to reproduce specific experimental data. For these reasons, PM6 is a fast method, however, it may lack in accuracy. *Ab initio* Hartree-Fock (HF) formalization employs orbitals to solve the electronic wave function without exchange-correlation effects (Schüürmann, 2004; Enoch, 2010). Finally, the electron-correlation included density functional theory (DFT) describes the energy of the system directly from the electron density (Schüürmann, 2004; Enoch, 2010). HF and DFT allow a more accurate description of electronic structures. Consequently, the choice of the quantum chemical method will influence the molecular descriptors and, ultimately, the statistics of QSAR models.

Nevertheless, a comprehensive understanding of this influence is still missing to date. Although Schüürmann (2004) provided a review on quantum chemical descriptors in QSARs, and others reported values for few quantum chemical descriptors calculated at different levels (Pasha et al., 2005; Eroglu et al., 2007; Reenu and Vikas, 2015), noticeably, there is no extensive study on the comparative dependence of descriptors computed by widely utilized software DRAGON on geometry optimization methods. Regarding the performance of QSAR models, a few studies recommended the use of a particular level, such as PM6 (Puzyn et al., 2008), HF with the 6-31G basis set (Rinnan et al., 2010) and DFT (Eroglu et al.; 2007; Becke, 1993; Lee et al., 1988; Pandith and Islam, 2013), mainly for quantum chemical descriptors (Puzyn et al., 2008; Pandith and Islam, 2013). Despite the interest in employing DRAGON descriptors, there are only a few studies contemplating the effect of using descriptors calculated by different quantum chemical methods on

the quality of QSAR models (Rinnan et al., 2010; Rybinska et al., 2016). In this thesis, a rational approach towards selecting an optimal geometry optimization method for improved QSAR/QSTR modeling was presented.

2.3.4. Training/Test Set Division

The dataset is divided into training set and test set to construct and internally and externally validate QSTR models, respectively. The training set should cover the test set and the test set should be representative of the training set. Hence, the dataset division should satisfy the following conditions: i) Each test set compound in the multi-dimensional descriptor space is close to those of the training set. ii) Each training set compound is close to those of the test set. iii) The representative points of the training set are distributed within the entirety of the dataset. It is accepted that the test set should have at least five compounds (Golbraikh et al., 2003). There are different division methods, such as periodical division and division based on Principal Component Analysis (PCA). In periodical division, the dataset is ranked in ascending (or descending) order of the dependent variable and starting from the second compound, for instance, every fifth or sixth compound is selected as test set, depending on the size of the dataset. Likewise, for PCA-based division, the dataset is ordered with respect to the molecular descriptors' first axis principal component (PC1) score (Jackson, 1991). For both divisions, compounds having the minimum and maximum value of the ordered toxicity or PC1 score are selected as training to ensure that the training set covers the entire range of both the dependent and independent variables.

2.3.5. Descriptor Selection

At the present, over 5,000 molecular descriptors can be calculated with the help of the DRAGON software. Moreover, a number of additional descriptors can be derived using the thermodynamic, physicochemical and quantum chemical parameters generated by the SPARTAN software. However, a valid QSTR model is anticipated to be as simple as possible (OECD, 2007). Therefore, in order to construct a mathematical model between the most relevant descriptors and toxicity, a selection procedure by means of different techniques should be applied. Often, an unsupervised variable reduction is applied to filter the constant (> 80%) and highly intercorrelated descriptors (pair-wise correlations among all pairs of descriptors, > 95%) due to their statistical insignificance previous to supervised variable selection. The term "unsupervised" denotes that there is no "supervision" of the response variables because they are not involved in analyzing the relationship between descriptors. Among the supervised variable selection techniques, All Subsets

(AS) and Genetic Algorithms (GAs) are the ones with common use (Gramatica et al., 2013). In a supervised approach, the response variables are used with all combinations of descriptors to construct the models by means of trial and error learning. AS is based on the generation of all the possible combinations of the available descriptors. This technique ensures that the best subset of variables is found, however, it is often not feasible due to the enormous number of combinations. GAs are one of the techniques used to overcome this issue. GAs are adaptive heuristic search algorithms inspired by the Darwin's notion of "survival of the fittest". As such, they provide an intelligent exploitation to solve the optimization problem, where the best solutions replace the less performing. GAs aim at finding an optimal subset of descriptors by maximizing (or minimizing) a selected fitness function, such as leave-one-out cross validation (Q^2_{LOO}). The use of the heuristic organized operations of "reproduction", "crossing", and "mutation" from random or user-defined starting "populations" generates the new "chromosomes", i.e., descriptor subsets. Tournament selection is a method for selecting a subset of descriptor in a GA (Gramatica et al., 2013). This method involves several "tournaments" run at random among the descriptors to fish out the best ones. The winner of each tournament is then selected for crossing, which allows the mixing of best descriptors coming from each parent. Based on the performance of these selected descriptors measured by the fitness function, the best descriptors are determined. However, for its random nature, GA can miss some relevant variables. In this study, in order to gain an insight into the best descriptors encoding the toxicity and to avoid a completely random start of the GA, AS and GA are used in combination to select the best descriptors out of all the possible best subsets of descriptors.

2.3.6. Construction of QSTR Models

The second principle of OECD validation criteria states that a QSTR model should have *"an unambiguous algorithm"*. A variety of algorithms based on linear or non-linear methods can be used to construct a mathematical model between the toxicity and selected descriptors. Among the linear regression approaches, MLR, based on ordinary least squares (OLS) is frequently used owing to its reasonable implementation and interpretability. In a linear regression model, the linear function between the dependent (response) and independent variables (descriptors) takes the form:

$$y_i = b_0 + \sum_{j=1}^{n} b_j x_{ij} + e_i \quad (2.2)$$

where y_i is the response, b_0 is the intercept, b_j are the coefficients to be estimated, x_{ij} are the values of the selected descriptors, and e_i is the random error (i.e. model residual). A positive coefficient of

x_{ij} suggests a positive contribution to toxicity, likewise, a negative coefficient implies a negative contribution. It should be noted that each coefficient should be significant at $p < 0.05$ level, which can be tested by a *t*-test (Roy et al., 2015a). The same equation rewritten in the matrix notation is as follows:

$$y = X\,b + e \tag{2.3}$$

where y is the vector of the observed responses, X is the matrix of the model where the columns are the descriptors, b is the vector of the coefficients, and e is the vector of errors corresponding to the vector of discrepancy between the observed and the predicted responses. The vector that estimates b vector of the coefficients is as follows:

$$\hat{b} = (X^T X)^{-1} X^T y \tag{2.4}$$

where X^T is the transposed X matrix and $^{-1}$ is the inverse matrix operation. It is important to note that, $(X^T X)^{-1} X^T$applies only if there is no multicollinearity between the descriptors. Q^2 Under Influence of *K* (*QUIK*) rule is a test measuring collinearity between variables in linear models (Todeschini et al., 1999). *QUIK* rule tests whether the total correlation among the block of descriptors (K_{XX}) is higher than the correlation among them and the responses (K_{XY}). The descriptors are regarded as not collinear if the condition of $(K_{XY}) - (K_{XX}) > \delta_K$ is met, where δ_K is a pre-defined threshold.

The variance-covariance matrix of $\hat{b}$ in Equation 2.4 is calculated to derive the standard error as follows:

$$\mathrm{Var}\left[\left(\hat{b} \mid X\right)\right] = \sigma^2 (X^T X)^{-1} \tag{2.5}$$

The standard error (i.e. the estimated standard deviation) of each coefficient $\hat{b}_j$ is equal to the square root of the *j*-th diagonal element of the variance-covariance matrix of $\hat{b}$:

$$\widehat{\mathrm{s.e.}}\,(\hat{b}_j) = \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}} \tag{2.6}$$

Use of excessive numbers of descriptors in a linear QSTR must be avoided to eliminate chance correlations. Therefore, the criterion of "Topliss and Castello rule" should be met for valid QSTR models. This rule states that the number compounds in the training set should be at least five times higher than the number of descriptors in the regression equation (Topliss and Castello, 1972).

When it is taken into account the non-linear nature of toxicity, an attempt will be important to model the toxicity by means of non-linear methods. Artificial Neural Networks (ANN) are one of the well-accepted non-linear techniques. ANN are computational tools comprising a group of processing elements (neurons) organized in subgroups (layers). Each subgroup can make its independent computations and pass the results to yet another subgroup. The last subgroup of one or more processing elements determines the output from the network. Practically, at a very simplified level, ANN mimic the way the human brain organizes and processes information, and the way the meaningful part of that information is identified and stored for future purposes. ANN typically start out with randomized weights for all their neurons, implying that they do not "know" anything in the beginning and should be trained for solving a particular problem for which they are intended. Training is the process of determining the weights, which are the key elements of ANN. The training process is usually as follows: first, examples of the training set are entered into the input nodes. The activation values of the input nodes are weighted and accumulated at each node in the first hidden layer. The total is then transformed by an activation function into the node's activation value. It in turn becomes an input into the nodes in the next layer, until eventually, the output activation values are found. The training algorithm is used to find the weights that minimize some overall error measure, such as the sum of the squared errors (*SSE*) or mean squared error (*MSE*).

CPANN method is a type of ANN often used in QSTR applications (Ertürk et al., 2012). It is a generalization of self-organizing maps technique, which is a mapping from multi-dimensional descriptor space onto 2D map of neurons through a non-linear projection. CPANN have two layers of neurons, the input (Kohonen layer) and the output layer arranged in 2D rectangular matrices. The Kohonen layer collects the 3D input variables and converts them into 2D map such that similar structures (i.e. similar descriptors) are located in the same neuron. It is an iterative procedure with many learning processes called epochs. During the learning process, the target values (toxicity) are given to the output layer, which has the same topological arrangement of neurons as the Kohonen layer. The algorithm then selects the neuron whose weights are the closest to the input values. The chosen neuron is referred to as the winning neuron. On the contrary, learning in the output layer is different. The position of objects is projected to the output layer. In the following step, the weights in the output layer are corrected in a way that they fit the output values (toxicities) of the

corresponding objects (Zupan et al., 1997). Finally, the trained network is then used for predictions. A typical architecture of CPANN is represented in Figure 2.1 (Figure adapted from Fjodorova et al., 2010).

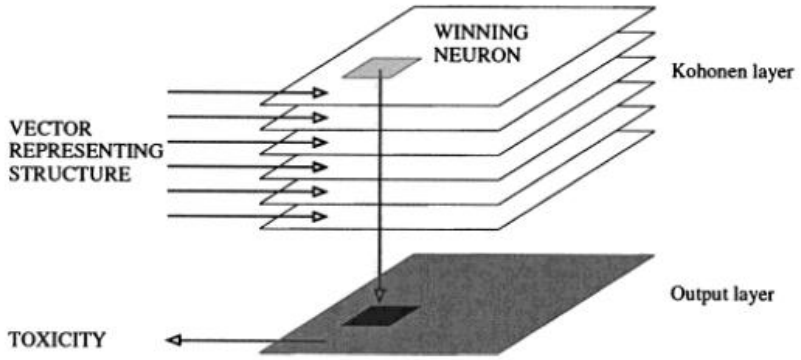

Figure 2.1. A typical architecture of CPANN.

In this thesis, OLS-based MLR method was used for all the linear models and CPANN method was used for non-linear modeling of algae data.

2.3.7. Applicability Domain

The third principle of OECD validation criteria states that a QSTR model should have *"a defined domain of applicability"*. A QSTR model is only applicable within its defined domain and can make predictions with certain reliability. The applicability domain (AD) of a QSTR model is the response (toxicity) and structural (descriptors) space, characterized by the properties of the training set. The predictive power of a QSTR model can be significantly changed due to the presence of outliers. Defining the AD highlights the outliers of the model, which consequently allows identifying the chemicals that are significantly different from the rest of the training set (structural outlier) and with less accurately predicted toxicity values (response outlier). Outlier detection could be functional also for the interpretation of the mechanism of action of the chemical and serves for the fifth principle of OECD validation criteria; *"a mechanistic interpretation, if possible"*.

Response outliers in MLR and CPANN models are determined if the predicted toxicity value is higher than ±3.0 standardized residuals, because the data points within ±3.0 standard deviations from the mean cover 99% of the normally distributed data. Different approaches can be used to define the structural AD of QSTR models. Among them, distance-based methods, such as the

leverage (OECD, 2007) for MLR models and the Euclidean distance (Minovski et al., 2013) for CPANN models are commonly used. The standardization approach (Roy et al., 2015b) is another method used for determining the AD.

The leverage of a chemical provides a measure of the distance of a chemical from the centroid of the model's training set. The values of model descriptors are employed to calculate the leverage matrix (or hat matrix) using the following formula:

$$H = X(X^T X)^{-1} X^T \tag{2.7}$$

where X is the matrix of model descriptors. The diagonal elements, h_{ii}, are the leverage values associated with each chemicals in the training set and ranges from 0 to 1. The larger the leverage the more distant a chemical from the center of the model. A threshold leverage, referred to as the critical hat value h^* is set at $3(p+1)/n$, where p is the number of descriptors appearing in the model and n is the number of compounds in the training set. Thus, a compound is considered structurally influential and can be identified as a high-leverage compound if the hat value of that compound is greater than h^*. For a visual definition, the standardized residuals are plotted against the leverages (Williams plot) to detect the outliers in both the descriptor and the response spaces. In a similar manner, the Insubria graph of calculated/predicted toxicity values versus diagonal hat values (Gramatica et al., 2013; Gramatica et al., 2014) is used to visualize the interpolated ($h < h^*$ condition, chemical falls in the structural AD of the training set) and extrapolated ($h > h^*$ condition, chemical falls outside the structural AD) predictions for the test set and external set chemicals. In this case, the response AD is the prediction range of the training set of the model.

The Euclidean distance (ED) is the ordinary distance, i.e., the line connecting two points defined by their 2D coordinates. The ED between the points A and B is calculated according to the Pythagorean formula:

$$ED_{A\text{-}B} = \sqrt{(X_{B1} - X_{A1})^2 - (X_{B2} - X_{A2})^2} \tag{2.8}$$

However, the CPANN method used for QSTR modeling deals with a multi-dimensional space of variables, the compound is defined as a vector and its ED is calculated to the centroid of the model's training set. A threshold ED, referred to as the critical ED value (ED^*) is therefore the

maximal distance in the training set of chemicals and defines the structural boundary condition for the AD (Minovski et al., 2013).

For the standardization approach a compound is identified as an outlier for the training set and outside the AD for the test set if; $[S_i]_{max(k)} > 3$ and $[S_i]_{min(k)} > 3$ or $[S_i]_{max(k)} > 3$, $[S_i]_{min(k)} < 3$ and $S_{new(k)} > 3$. $[S_i]_{max(k)}$ is the maximum standardized value of the descriptor *i* for the compound *k*, $[S_i]_{min(k)}$ is the minimum standardized value of the descriptor *i* for the compound *k*, and the last term is $S_{new(k)} = \bar{S}_k + 1.28\sigma_{S_k}$, where $\bar{S}_k$ is the mean of standardized values of the descriptors for the compound *k* and σ_{S_k} is the standard deviation of the standardized values of the descriptors for the compound *k* (Roy et al., 2015b).

2.3.8. Statistical Quality and Validation of QSTR Models

Evaluation of the statistical quality and validation is a series of procedures used to measure the goodness-of-fit, robustness, reliability, and predictivity of a mathematical model. The fourth principle of OECD validation criteria requires any QSAR model to have *"appropriate measures of goodness-of-fit, robustness, and predictivity"*. Therefore, any QSTR model should be validated internally and externally (OECD, 2007). Internal validation is applied on the training set and model itself, whereas external validation covers testing the model on a test set. To date, the modern QSAR literature has collaboratively defined rigorous and sophisticated metrics and criteria for these purposes (Gramatica and Sangion, 2016).

Goodness-of-fit is measured by the coefficient of determination (R^2), commonly calculated for evaluating how good the model's ability to reproduce the data that is used for the model development:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} \quad (2.9)$$

where y_i and $\hat{y}_i$ are the observed (experimental) and calculated toxicity of the compound *i*, respectively, $\bar{y}$ is the mean of the observed toxicity, *RSS* is the residual sum of squares and *TSS* is the total sum of squares. In general, a QSTR model is considered acceptable if $R^2 > 0.60$ (Golbraikh and Tropsha, 2002).

Robustness is related to the sensitivity of the model parameters to changes in the training data, thus, the internal performance of the model, which is measured by the leave-one-out (Q^2_{LOO}) or leave-many-out (Q^2_{LMO}) cross-validation. The following formula is calculated as an internal validation metric when iterative cross-validation is applied to verify the stability, i.e., robustness of a QSTR model:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{PRESS}{TSS} \quad (2.10)$$

where y_i and $\hat{y}_i$ are the observed and predicted toxicity of the compound *i* by the model when the object *i* is not in the training set, respectively, $\bar{y}$ is the mean of the observed toxicity, and *PRESS* is the predicted residual sum of squares. In general, Q^2 and Q^2_{LOO} are used interchangeably. Exclusion of multiple objects in each iteration is the leave-many-out cross-validation procedure and yields the Q^2_{LMO} parameter. A QSTR model is considered robust if $Q^2 > 0.50$ (Golbraikh and Tropsha, 2002).

Reliability of a QSTR model is determined by Y-scrambling procedure, which identifies a possible chance correlation between the toxicity and descriptors. In a procedure referred to as response randomization, the toxicity variables of the training set are shuffled in such a way that each object in the descriptors vector is no longer associated with its correct toxicity. Following the procedure many times, new coefficient of determination (R^2_{Yscr}) and cross-validation (Q^2_{Yscr}) metrics are calculated. Significantly low values of the parameters indicate that the original model was not built by chance correlation

Predictivity refers to the accuracy of the predictions provided by the model for compounds not used in the model development, therefore, is about the external performance. An external Q^2 function proposed by Shi et al. (2001); Q^2_{F1} is one of the first validation parameters assessing the predictive power of a model:

$$Q^2_{F1} = 1 - \frac{\sum_{i=1}^{n_{test}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{test}} (y_i - \bar{y}_{tr})^2} \quad (2.11)$$

where the *TSS* of the test set is calculated using the training set mean. In 2008, Schüürmann and co-workers (2008) reported a new metric; Q^2_{F2} demonstrating that an increasing difference between $\bar{y}_{tr}$ and $\bar{y}_{test}$ may result in on overestimation of the model's predictive ability:

$$Q^2_{F2} = 1 - \frac{\sum_{i=1}^{n_{test}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{test}} (y_i - \bar{y}_{test})^2} \quad (2.12)$$

where, on the contrary, the *TSS* of the test set is calculated using the test set mean. Following to this, Consonni and colleagues (2009) proposed a new parameter; Q^2_{F3} to compensate the drawbacks of earlier Q^2 functions:

$$Q^2_{F3} = 1 - \frac{\left[\sum_{i=1}^{n_{test}} (y_i - \hat{y}_i)^2\right]/n_{test}}{\left[\sum_{i=1}^{n_{tr}} (y_i - \bar{y}_{tr})^2\right]/n_{tr}} \quad (2.13)$$

where n_{test} and n_{tr} are the number of compounds in the test and training set, respectively. Including n_{test} and n_{tr} in the equation makes this parameter independent of the distribution and size of the test set. The predictivity of a QSTR model is considered acceptable when all three external Q^2 functions are higher than 0.60 (Chirico and Gramatica, 2011). The concordance correlation coefficient (*CCC*) of Lin (1989; 1992) was later adapted to QSTR models by Chirico and Gramatica (2011; 2012) to assess the agreement between the observed and predicted values:

$$CCC = \frac{2\sum_{i=1}^{n_{test}} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{test}} (y_i - \bar{y})^2 + \sum_{i=1}^{n_{test}} (\hat{y}_i - \bar{\hat{y}})^2 + n_{test}(\bar{y} - \bar{\hat{y}})^2} \quad (2.14)$$

where $\bar{\hat{y}}$ is the mean of the predicted toxicity. A QSTR model is considered acceptable if *CCC* > 0.85 (Chirico and Gramatica, 2011). In addition, r^2_m average and Δr^2_m are calculated to evaluate the performance of models:

$$r^2_m = r^2\left(1 - \sqrt{r^2 - r^2_0}\right) \text{ and } \Delta r^2_m = \left|r^2_m - r'^2_m\right| \quad (2.15)$$

where r^2 is the coefficient of determination for test set, r^2_0 is the r^2 without intercept, and r'^2_m is coefficient of determination for observed toxicity on the x-axis vs. predicted toxicity on the y-axis. The acceptance thresholds for these metrics are $r^2_m > 0.50$ and $\Delta r^2_m < 0.20$ (Ojha et al., 2011).

Moreover, based on Golbraikh and Tropsha (2002) criteria, QSTR models are considered to have acceptable prediction power, if they meet the following conditions in addition to the above criteria:

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

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

$\left|R^2 - R_0'^2\right| < 0.3$ (2.16)

where R_0^2 (predicted vs. experimental) and $R_0'^2$ (experimantal vs. predicted) are coefficients of determination without intercept, k and k' are the corresponding slopes.

MAE_{test} is the mean absolute error calculated over the 95% of the test set data when $n_{test} > 10$ (Roy et al., 2016). Regarding this approach, $MAE_{test} \leq 0.1 \times$ training set range (TSR) and $MAE_{test} + 3 \times \sigma \leq 0.2 \times$ TSR conditions should be met for good predictions.

2.4. QSTRs in Aquatic Toxicology: A Brief History and Current Status

Studies on the relationship between chemical structure and toxicity have been documented since the end of the 19th century. In 1893, Richet (1893) stated one of the basic principles of toxicity for alcohols and ethers: *"the more soluble they are the less toxic they are"*. Six years later, Meyer (1899) and Overton (1899) described the correlation between the lipid-water partition coefficient and narcotic activity of many substances, which is still one of the principle determinants of toxic action. According to Ferguson (1939), the important criterium that determines the toxicity of a substance in fish is not the concentration but the chemical potential. Narcosis follows Ferguson's principle: the rate-limiting step for narcosis is the ability of the chemical to reach the site of action.

After a series of work by Hansch and collaborators (Hansch et al., 1962; Hansch et al., 1964; Hansch and Glave, 1971), the investigation of structure-activity relations has begun to draw attention. Hansch used the following empirical equation describing one of the first QSARs as a function of electronic, hydrophobic, and steric parameters:

$$\log 1/C = k_1 \log P - k_2 (\log P)^2 + k_3 pK_a + k_4 E_s + \ldots + k_5 \quad (2.17)$$

where C is the concentration of a substance inducing a certain biological response, P is a partition coefficent, usually in n-octanol/water system ($\log K_{ow}$), K_a is acid-dissociation constant, E_s is a steric parameter, and k_n are the coefficients for fitting.

According to McFarland (1970), toxicity is the combined result of penetration of the toxicant into the biological membrane and the interaction with the site of action. McFarland's approach is represented mathematically as follows:

$$\log\ 1/\text{toxicity} = \text{A}\ \log\ \text{penetration} + \text{B}\ \log\ \text{interaction} + \text{C} \qquad (2.18)$$

This approach is toxicokinetics and toxicodynamics based and does not involve electronic and steric effects. While the toxicokinetics phase includes chemical uptake and metabolic activation, the toxicodynamics phase involves the molecular and biochemical responses, leading to cellular or physiological responses.

Könemann (1981) studied the LC_{50} of industrial pollutants towards the guppy, *Poecilia reticulata*, and found that the octanol-water partition coefficient (log K_{ow}) is the only variable explaining the toxicity. The QSTR model yielded good estimations for the most of the studied compounds with log $K_{ow} < 6$. This study has led to the emergence of the concept of "baseline toxicity". Baseline toxicity (also known as narcosis) is the minimal toxic effect, which is due only to the lipophilicity measured as log K_{ow}. Chemicals showing baseline toxicity are classified as inert or narcotic chemicals. Narcotic chemicals cause only non-covalent and reversible alterations at the site of action. All organic chemicals have the potential to cause narcosis. Narcosis can be polar or non-polar depending on the electronic properties of the chemical.

Afterward, Verhaar and co-workers (1992) established a classification scheme based on structural alerts to assign organic chemicals to one of four categories as follows:

1. *Inert chemicals (non-polar narcotics)* that are not reactive and do not interact with specific receptors. Inert chemicals exert narcosis-type baseline toxicity, i.e., the toxic potency depends solely on lipophilicity.
2. *Less inert chemicals (polar narcotics)* are not reactive but slightly more toxic than that of inert chemicals. Higher toxicity is due to the polar moieties.
3. *Reactive chemicals* show significantly higher toxicity than baseline toxicity (excess toxicity). Such chemicals form irreversible covalent bonds with amino acid protein residues. In this context the term "reactive" encompasses a wide spectrum of competitive electro- and nucleophilic, redox, and free radical processes.

4. *Specifically acting chemicals* react with specific biological targets through a different mode of action (MoA).

The contemporary QSTR research owes much to these pioneering studies. Furthermore, development of the theoretical molecular descriptors (Todeschini and Consonni, 2009) and progresses achieved in informatics and the power of computers have enabled the development of a large number of QSTRs for the prediction of a plenty of toxicity endpoints, using various statistical techniques.

The current status and future needs of QSTRs in environmental toxicity predictions and their regulatory uses were summarized in a perspective by Cronin (2017). Although a plenty of QSTR models are reported in the literature, remarkably fewer QSARs have been developed on algae compared to other aquatic species such as fish (Tugcu et al., 2012; Sangion and Gramatica, 2016) and *Daphnia* (Sangion and Gramatica, 2016; Aalizadeh et al., 2017). A selection of relevant QSTR models for algae is discussed in the following paragraph. Particular focus was given to the models that are comparable (based on species and testing duration, chemical diversity, and modeling approaches) to the developed models under the scope of this thesis.

Aruoja et al. (2011) modeled the toxicity data of a congeneric set of anilines and phenols to algae *P. subcapitata* in order to support the hazard classification for REACH. Comparison of the experimental toxicity data with the predictions made using the existing QSAR model suggested that the toxicity of phenols to algae might be modeled with a simple hydrophobicity-based equation. However, aniline toxicity to algae depended on other characteristics in addition to the log K_{ow}. Gramatica et al. (2012) modeled the toxicity of a small dataset of (benzo-)triazoles to *P. subcapitata* with molecular descriptors other than log K_{ow} for screening and prioritization of chemicals. Chen et al. (2012) reported a model correlating the toxicity of a small dataset of propargylic alcohols to *P. subcapitata* and log K_{ow} for risk assessment purpose. Ertürk and Saçan (2013) analyzed a novel dataset of 30 phenols to *Chlorella vulgaris* and reported that the toxicity of polar narcotics and respiratory uncouplers are correlated to the pH-corrected hydrophobicity parameter, demonstrating the importance of ionization in the *C. vulgaris* test system. Despite this study generated toxicity data in a single laboratory, the model was built with a limited number of chemicals. In another study, Aruoja et al. (2014) studied the 72-h experimental toxicity data of 108 non-polar and polar narcotics towards *P. subcapitata*. Expectedly, toxicity of the non-polar chemicals well correlated with the log K_{ow}, providing a baseline model. Pramanik and Roy (2014) modeled a diverse set of 74 organic chemicals to *P. subcapitata* with five descriptors. Sangion and Gramatica (2016) reported a

4-descriptor MLR model for pharmaceuticals to predict the toxicity of *P. subcapitata*. Interestingly, log K_{ow} did not appear in this model. Lately, Tugcu et al. (2017) developed a linear model for the prediction of 96-h *C. vulgaris* toxicity using phenols. Despite their obvious contribution to the QSTR literature, these models were built using relatively small datasets, thus, providing narrower ADs. Global QSTR models for algal toxicity predictions based on a wide range of chemicals (Singh et al., 2014; Villain et al., 2014) and multispecies (Basant et al., 2015) are notably few and do not fully fulfill the OECD validation requirements (OECD, 2007). This thesis provides linear and non-linear validated QSTR models with significantly wider ADs using the toxicity data of structurally diverse chemicals towards mixed algae species.

The existing QSTR reports on the rainbow trout and *D. magna* and the interspecies QTTR literature are outlined in the following paragraphs.

There are studies on the toxicity prediction of the rainbow trout (*O. mykiss*) for organic chemicals including pharmaceuticals (Tugcu et al., 2012; Cassani et al., 2013; Sangion and Gramatica, 2016), however, there is only one report on the toxicity of PPCPs to the RTL-W1 cell line (Schnell et al., 2009). In this study, the log K_{ow} and cytotoxicity of the chemicals had no relationship. Therefore, it was worthwhile to search for descriptors that can be used to relate cytotoxicity to chemical structures. In this thesis, a validated QSTR model with a wide AD was presented for cytotoxicity prediction of RTL-W1 cell line for the first time.

QSTR modeling of *D. magna* has been extensively studied (Sangion and Gramatica, 2016; Aalizadeh et al., 2017). Nevertheless, emerging demands for a more holistic approach to regulatory environmental safety assessment require the results of more diverse aquatic invertebrates besides *D. magna* (Brown et al., 2016). Recently, *D. japonica* has been proposed as an ideal alternative to *Daphnia* due to its unique properties for environmental toxicology studies (Li, 2008; Li, 2012a; Li, 2012b; Li, 2013a; Li, 2013b). Therefore, it was noteworthy to study the toxicity modeling of *D. japonica*.

Kar et al. (2016) provided a comprehensive review of earlier efforts towards *Daphnia*-fish QTTR modeling (Kar and Roy, 2010; Cassani et al., 2013; Furuhama et al., 2015), as well as the web-based Interspecies Correlation Estimation (ICE) application (Dyer et al., 2006; US EPA, 2016) of the United States Environmental Protection Agency (US EPA).

Despite the critical ecological importance of *D. japonica*, interestingly, there is no report on the relationship between toxicity and the log K_{ow}, QSTR and QTTR modeling of *D. japonica*. This thesis provides the first and validated QSTR model with a wide AD for the acute toxicity prediction to *D. japonica*. Likewise, interspecies quantitative toxicity relationship between *D. magna* and *D. japonica* was investigated and a validated QTTR model was presented in this thesis, for the first time.

3. MATERIALS AND METHODS

3.1. Experimental Datasets

All the experimental datasets were evaluated in terms of their modelability in accordance with the characteristics outlined in Section 2.3.1. Whenever needed, the datasets were cleaned and curated prior to modeling. All the experimental toxicity variables were transformed to negative logarithm scales for modeling purpose, yielding either pEC_{50} or pLC_{50} values.

3.1.1. Case Study Dataset

A case study dataset was used in order to investigate the impact of different quantum chemical methods employed for geometry optimization on the molecular descriptors and statistical quality of QSTR models. This dataset comprising the experimental binding affinity data of 88 drug and drug-like molecules to human serum albumin (HSA) were taken from Colmenarejo et al. (2001) and presented in Table A.1. Binding affinity was measured as the binding constant obtained from the retention time on an immobilized HSA column using affinity chromatography (log K_{HSA}). Seven compounds were excluded from the original data of 95 compounds for a variety of reasons (captopril: experimental error; ebselen, minocycline and sancycline: geometry optimization failure due to self-consistent field convergence problem; digitoxin: geometry optimization termination failure due to high molecular weight; verapamyl: name did not match with a reliable CAS number; and zidovudine: due to negatively charged azide moiety). The experimental data are normally distributed (Colmenarejo et al., 2001) and range from –1.39 to 1.34 (Table A.1).

3.1.2. Algae Dataset

Experimental 72-h mixed freshwater algae toxicity data (pEC_{50} in mol/L unit) were extracted from Fu et al. (2015) and meticulously curated as outlined in Section 2.3.1. The initial acute toxicity data ranged from 0.23 to 9.66 for 518 structurally diverse chemicals. Charged and salt molecules were removed from the compiled dataset for modeling purpose. The correctness of the CAS numbers and corresponding chemical structures for the remaining 490 chemicals were checked. Great majority of the

data consisting of 250 different functional classes belonged to the OECD recommended species (84.5% *P. subcapitata* and 7.6% *Desmodesmus subspicatus*) (OECD, 2006). The diversity in chemicals and algae are shown in Figure 3.1. Dataset was grouped based on the reported MoA (Fu et al., 2015). Group 1 contained 61 non-polar narcotics (chemicals with baseline toxicity) and Group 2 contained 74 polar narcotics (less inert chemicals). Group 3 comprised 490 entities covering Group 1 and 2 and the remaining chemicals with unknown MoA. Since some of the chemicals in the dataset had more than one experimental endpoint, the following rules were applied to determine a single pEC_{50} value: i) Arithmetic mean was accepted when the calculation methods were the same for all endpoints of the same species (mostly for *P. subcapitata*). ii) Value with a known calculation method was accepted when the calculation method was not available for the other reported value of the same species. iii) If available, the value obtained for *P. subcapitata*, if not, arithmetic mean was accepted for the same chemical having different endpoint values for different algal species. After obtaining single endpoints, the possible cliffs in the response and descriptor spaces were checked and the Group 3 was further filtrated in accordance with Section 2.3.1, resulting in 455 compounds in the final dataset. Group 3's toxicity data appeared to follow a normal distribution (Kolmogorov-Smirnov test ($p > 0.05$)) and ranged from 1.21 to 8.04. Chemicals and pEC_{50} values are presented in Table B.1. Excluded chemicals, their toxicity and descriptor values are available as supplementary material (SM) online at the article's web address https://onlinelibrary.wiley.com/doi/abs/10.1002/etc.3620 (Önlü and Saçan, 2017b). Of the 455 chemicals in the dataset, around 54% were designated as HPV chemicals (Table B.1).

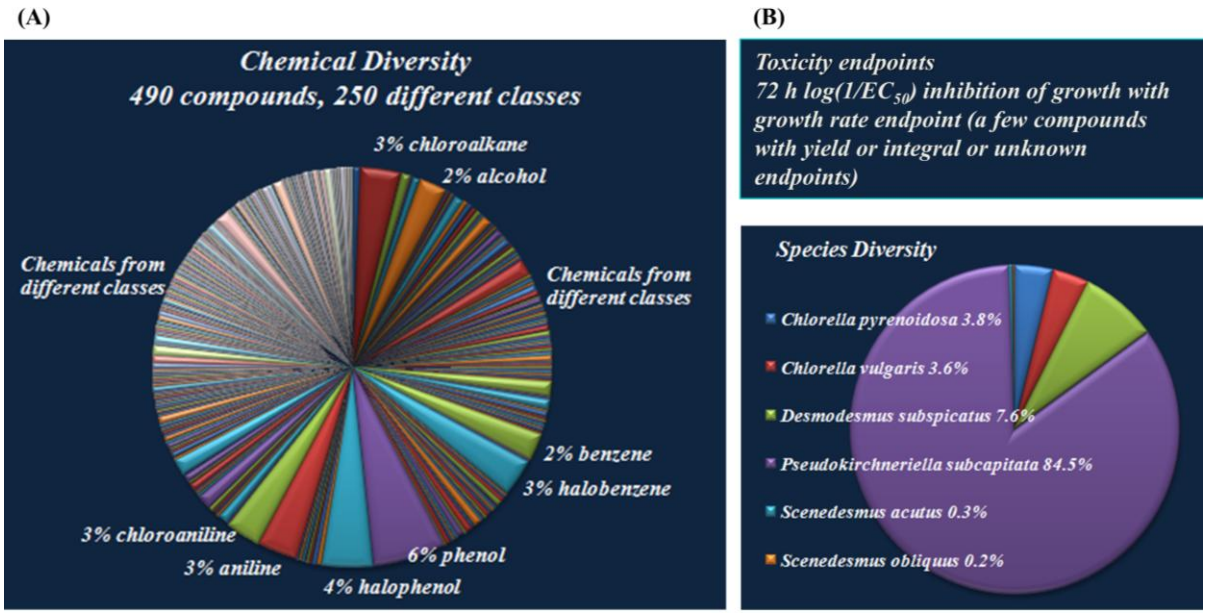

Figure 3.1. Diversity in chemicals (A) and algae species (B).

3.1.3. RTL-W1 Cytotoxicity Dataset

Experimental RTL-W1 cytotoxicity data were taken from Schnell et al. (2009) including 11 pharmaceuticals from different therapeutic classes (anti-inflammatory drugs: diclofenac, ibuprofen, ketoprofen, naproxen, fibrates: bezofibrate, clofibrate, fenofibrate, gemfibrozil; antidepressives: fluoxetine, fluvoxamine, paroxetine), as well as five synthetic musks from two major groups; nitro musks (musk ketone and musk xylene) and polycyclic musks (celestolide, galaxolide, tonalide). Because cytotoxicity data on RTL-W1 is scarce, additional data for two industrial chemicals; acid blue 80 (an anthracenedione dye) and pentachlorophenol taken from Tee et al. (2011) and Schreer et al. (2005), respectively, were also included to increase the toxicity range. The pEC_{50} values obtained from the AB assay ranged from –2.44 to –0.84 (in μmol/L unit). The pEC_{50} values obtained from the CFDA-AM assay ranged from –2.43 to –0.83 (in μmol/L unit). Both datasets appeared to follow a normal distribution (Kolmogorov-Smirnov test ($p > 0.05$)).

3.1.4. *Dugesia japonica* Dataset

Experimental data for 55 CEC from different chemical classes, such as PPCPs, EDCs including synthetic and natural hormones, insecticides, pesticides, and nonionic surfactants, were compiled from the literature (Li, 2008; Li, 2012a; Li, 2012b; Li, 2013a; Li, 2013b; Hagstrom et al., 2015). The reported 48-h pLC_{50} (in μmol/L unit) appeared to follow a normal distribution (Kurtosis = 0.16, skewness = −0.56) and ranged from −5.14 to −0.42. Of the 55 chemicals in the dataset, 24% were designated as HPV chemicals. The normal distribution plots and HPV status of each chemical in the dataset are available as SM online at the article's web page https://doi.org/10.1016/j.jhazmat.2018.02.046 (Önlü and Saçan, 2018).

3.2. Structures, Molecular Descriptors and Modeling Datasets

All the quantum chemical calculations were carried out with the SPARTAN software (SPARTAN v. 10, 2011). To this end, all the chemical structures were drawn manually and the lowest energy conformer of each molecule was further geometry optimized using a certain geometry optimization method. Vibrational analyses on the optimized geometries were performed to verify the absence of imaginary frequencies, ensuring that the geometry is a minimum energy point in the potential energy

surface rather than a transition state. Quantum chemical descriptors such as the dipole moment, the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}), gas-phase energy (E), and hardness (($E_{LUMO} - E_{HOMO}$)/2)) were calculated. The software DRAGON (DRAGON v. 6, 2013) was used to process the optimized geometry of each structure to generate descriptors. Finally, modeling datasets were prepared by merging the descriptors and the response variables.

3.2.1. Case Study Descriptors, Descriptors Analyses and Datasets

Each molecule in the dataset was geometry optimized using three methods. The methods were as follows: i) semi-empirical PM6 (Stewart, 2007), ii) *ab initio* HF with 6-31G(d,p) basis set (Petersson et al., 1988) and iii) DFT with the standard Becke's three-parameter exchange potential and the Lee-Yang-Parr correlation functional (B3LYP) (Becke, 1993; Lee et al., 1988) using 6-31G(d,p) basis set. 6-31G(d,p) was preferred as it provides an adequate compromise of speed and accuracy when hydrogen is the site of interest. It adds d-type polarization functions to heavy atoms as well as p-type polarization functions to hydrogen atoms, thus improves the modeling of core electrons and total energy of the system (Petersson et al., 1988). Molecular descriptors were generated using the optimized geometry of each structure obtained from three different levels. Thus, three separate data matrices each of comprising 3182 descriptors (a total of 9546) were obtained for 88 chemicals at each quantum chemical level for further analyses (Full data is available as SM online at the article's web page http://dx.doi.org/10.1080/1062936X.2017.1343253 (Önlü and Saçan, 2017a)).

Measures of skewness and kurtosis (Joanes and Gill, 1998) of each set of variables were calculated and used as evidence of normality. The majority of the data consists of continuous variables (82%). Based on these, Pearson correlation coefficients (r, significant at $p < 0.05$) for pairs of each set of variable were computed to investigate the quantitative influence of each quantum chemical method on the value of descriptors using the R software (R Development Core Team, 2012).

Finally, three modeling datasets were prepared for each quantum chemical level: Dataset 1 (DS1, all descriptors + log K_{HSA}), Dataset 2 (DS2, only affected/sensitive descriptors + log K_{HSA}), and Dataset 3 (DS3, all descriptors except the descriptors of the proposed model in the present study + log K_{HSA}). DS1 was used to build the best possible model to be proposed. DS2 was used to monitor the impact of

geometry optimization methods on the performances of QSAR models built only with the sensitive descriptors. DS3 was used to monitor the same impact with a broader comparison possibility.

3.2.2. QSTR Modeling Descriptors and Datasets

Based on the results of the case study, QSTR modeling datasets for algae, RTL-W1 cytotoxicity, and *D. japonica* were generated at the semi-empirical PM6 method following the general procedure outlined in Section 3.2. For the *D. japonica* dataset, in addition to the quantum chemical and DRAGON descriptors, the log K_{ow} and *D. magna* 48-h pEC_{50} (in µmol/L unit) data were also used for hydrophobicity and QTTR modeling, respectively. The log K_{ow} values were retrieved from the Danish (Q)SAR Database (http://qsar.food.dtu.dk). Preferably experimental, otherwise, estimated values (EPI WSKOW v1.42) of log K_{ow} were used. Likewise, *D. magna* 48-h pEC_{50} values were collected from the literature (Li, 2013b; Sangion and Gramatica, 2016; Aalizadeh et al., 2017). Of the 55 chemicals in the original dataset, 26 chemicals had a reported experimental pEC_{50}. Consequently, QSTR datasets were finally prepared by incorporating the experimental toxicity values for each aquatic species.

3.3. Modeling and Validation

An unsupervised variable reduction was applied as outlined in Section 2.3.5 to filter the constant (> 80%) and highly intercorrelated descriptors (pair-wise correlations among all pairs of descriptors, > 95%) due to their statistical insignificance prior to training/test set division, descriptor selection, and actual modeling. QTTR modeling dataset was exempted from this reduction since the 48-h pEC_{50} *D. magna* data was the only descriptor used. Descriptor reduction, training/test set division, descriptor selection and linear modeling were carried out using the software QSARINS 2.2.1 (Gramatica et al., 2013; Gramatica et al., 2014). Flowchart of QSTR modeling is presented in Figure 3.2.

3.3.1. Training/Test Set Division

All the modeling datasets were divided into training set (to develop models) and test set (to validate the developed models). Division was done in accordance with the procedures outlined in Section 2.3.4.

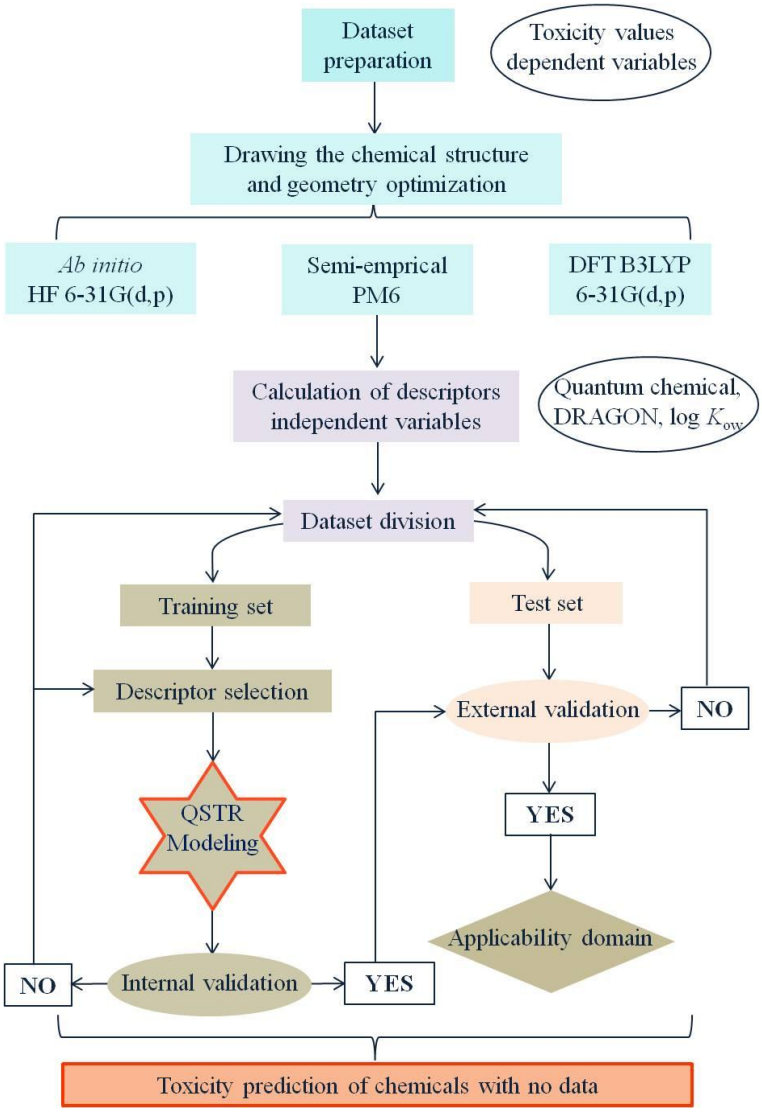

Figure 3.2. Flowchart of QSTR modeling.

Two different methods at varying ratios were used for the training set (~75-85%) and the test set (~15-25%) division in order to avoid a possible bias that might have arisen if a single approach was applied: i) by ordered response and ii) by ordering the molecules based on the molecular descriptors' PC1 score. For the QTTR modeling, only response-based division was applied. For each division, chemicals with the minimum and the maximum values of the ordered response or PC1 score were selected as the training set compound. This way, it was ensured that the training set covers the entire range of both the response and the descriptors spaces. Conseqently, the optimal training/test set division for each model was determined. The same division of the DS1 was also applied to DS2 and DS3 for a one to one comparison. Likewise, the same division was used for linear and non-linear modeling of the algal QSTR models for comparison purpose.

3.3.2. Descriptor Selection, Linear Modeling and Validation

Descriptor selection was performed over the modeling datasets using the AS and the GA methods as outlined in Section 2.3.5. MLR based on the OLS method was used for model development, as OLS offers an optimal solution given the redundancy in the dataset was previously mitigated. First, all low-dimensional models (up to 2–3 descriptors, corresponding to all combinations) were calculated using the AS facility to preliminarily explore the best descriptors encoding the response and to avoid a completely random start of the GA. The best subset of descriptors appeared at this step was used as the core of chromosomes of the initial population for the GA. Next, based on the tournament selection, GA was used to explore the solution space by maximizing the fitness function with population size = 100, mutation rate = 20 and number of generations = 1000. Q^2_{LOO} was chosen as the fitness function because it provides a measure of model stability and robustness. Since the descriptor selection was done by means of the GA, the evaluation of MLR based on OLS never happened on the whole dataset but smaller subsets selected by the GA. However, for its nature, GA may miss out some relevant variables. In order to handle this, the GA was run multiple times to ensure a broader exploration of the solution space. Running multiple nested instances of the GA, therefore, allowed to address a possible stability problem and to attenuate the risk of noise. Following these procedures repeatedly, a population of good models was generated. For interspecies toxicity relationship, simple linear regression allowed to construct one-parameter QTTR models using different training/test set divisions.

Fitting, robustness, and predictive performance of the models were rigorously evaluated in accordance with the up-to-date fit, internal, and external validation parameters and criteria explained in Section 2.3.8.

3.3.3. Non-linear Modeling

CPANN models were developed using the algae dataset to compare the performance of a non-linear approach on the data modeling. In an attempt to understand how the descriptors of the final MLR-based model would contribute to a non-linear model, CPANN models were built with the same descriptors. Likewise, the same final training/test set division used for the MLR-based model was used for comparison purpose. Training of the model was carried out with different dimensions of network architecture (19×19, 20×20, 21×21, 22×22, 23×23, and 24×24) and gradually increasing number of

epochs (from 50 to 1000 increasing by 50). Number of neurons in x and y-directions were set at 5. Minimal correction factor was set at 0.01. Maximum correction factor was set at 0.50. Model quality was assessed based on the higher values of R^2 for the training and test sets and Q^2_{LOO}. Consequently, a population of good models was generated. CPANN model development was performed using the modules developed at the Slovenian National Institute of Chemistry (Zupan and Gasteiger, 1999).

3.4. Applicability Domain Definition

The ADs of the MLR-based models were defined as outlined in Section 2.3.7. For the leverage approach, the standardized residuals of the response variables were mapped against the leverages (hat values) for a visual characterization of the AD (Williams plot). Chemicals exceeding certain threshold values were identified as response and structural outliers. Response threshold values were set at ±3.0 standardized residuals and indicated as horizontal dashed lines in the Williams plots. Leverage threshold was fixed at the critical hat value for each model and represented by a vertical solid line in the Williams plots. Furthermore, the Insubria graphs (Gramatica et al., 2013; Gramatica et al., 2014) of predicted toxicities against leverages/hat values were reported for the training/test set and the external set chemicals (i.e. chemicals with no experimental data) to visually examine the prediction applicability and structural coverage of the proposed models. The response-prediction applicability was indicated using horizontal dashed lines representing the prediction range of models in the graphs. Similarly, a vertical solid line representing the critical values was used to indicate the structural coverage of the models. Thus, data predicted for high leverage compounds was considered extrapolated by the model. For the definition of the AD of the CPANN model, Euclidean distances were used instead of leverage values.

3.5. Selection of the Best Model

The multi-criteria decision making (MCDM) procedure implemented in the QSARINS software ranks the model performances from 0 to 1, with 0 as the worst and 1 as the best, based on fitting, internal and external validation. Among the developed good MLR models, models with the best MCDM score, fulfilling the OECD validation requirements (OECD, 2007) as well as the statistical acceptance criteria (Section 2.3.8), and with the least possible number of descriptors were selected as the final models. In case of QTTR models, the last criterion was not evaluated due to irrelevance.

In case of the non-linear CPANN models, the model with the best compromise between R^2 and Q^2_{LOO}, as well as the number of epochs and network architecture was chosen as the final model. Consequently, the model with the highest R^2 and Q^2_{LOO} values and the lowest dimension of network architecture was chosen.

4. RESULTS AND DISCUSSION

4.1. Impact of Geometry Optimization Methods on QSAR Modeling: A Case Study

4.1.1. Geometry Optimization Method versus Molecular Descriptor Values

The dependence of the descriptors on the geometry optimization method was elucidated. Measures of skewness (mean = 1.06; σ = 1.75) and kurtosis (mean = 4.71; σ = 8.66) of each set of variable justified the applicability of a pair-wise Pearson analysis. Distribution of variable skewness kurtosis and the full results of Pearson correlations between pairs of descriptors are available online as SM at the article's web page http://dx.doi.org/10.1080/1062936X.2017.1343253 (Önlü and Saçan, 2017a). Of the 3182 descriptors, 1084 were found to be affected by the geometry optimization method (Figure 4.1), as evidenced by the Pearson correlation coefficients ($r < 1$). Molecular descriptors affected by the geometry optimization method were presented in Table C.1 together with their type and chemical meanings.

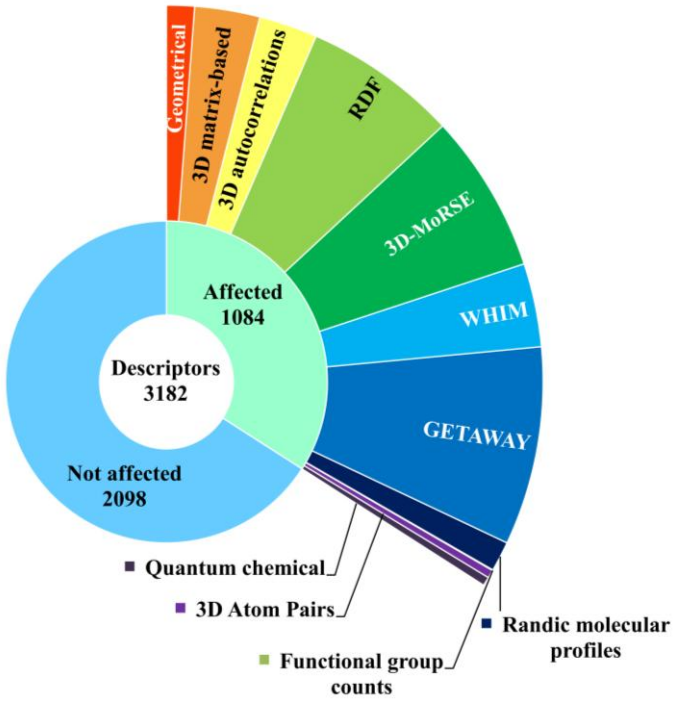

Figure 4.1. Sunburst plot of descriptor analyses.

Of the total affected descriptors, 3.5% were geometrical, 8.2% were 3D matrix-based, 7.4% were 3D autocorrelations, 19.4% were radial distribution function (RDF), and 20.0% were 3D-molecule representation of structures based on electron diffraction (3D-MoRSE). Likewise, 10.5% were weighted holistic invariant molecular (WHIM), 25.0% were geometry, topology, and atom-weights assembly (GETAWAY), 3.8% were Randic molecular profiles, 0.1% is functional group counts, 1.0% were 3D atom pairs, and 1.1% were quantum chemical.

Nine descriptor groups encoding 3D structures were found to be sensitive to the quantum chemical calculation method: geometrical, 3D matrix-based, 3D autocorrelations, RDF, WHIM, GETAWAY, Randic molecular profiles, 3D-MoRSE, 3D atom pairs and quantum chemical. Almost all of the 3D-MoRSE descriptors (217 out of 224) were influenced by the geometry optimization method. Noticeably, none of Mor01 (signal-01 3D-MoRSE descriptors, calculated based on the scattering parameter = 0 $Å^{-1}$) were affected by the methods considered in this study. Although 3D-MoRSE descriptors are dependent on 3D structures, given the mathematical background based on electron diffraction studies and EDs between atoms, assigning 0 $Å^{-1}$ to the scattering parameter in the formula makes the inter-atomic distance relevant term equals to one regardless of the different weights used, thus eliminates the impact of geometry (Devinyak et al., 2014). Remarkably, one 3D structure-independent functional group counts descriptor, which is also the only single intra-molecular parameter among all, nHBonds (number of intra-molecular H-bonds) was weakly influenced ($r > 0.9$) by the quantum chemical level.

Out of 1084 descriptors being affected by the quantum chemical method, the majority of correlations ranged between 0.75 and 1 for all three cases (89% for PM6 vs. HF, 95% for HF vs. DFT, and 93% for PM6 vs. DFT), indicating low sensitivity. Therefore, to obtain deeper insight into the significant effects of optimization methods on the descriptors, a specific focus was given to the descriptor groups with $r < 0.75$. Pearson correlation results and the number of descriptors significantly affected ($r < 0.75$) from each group for each pair of geometry optimization method are presented in Table C.2 and Table 4.1, respectively.

3D-MoRSE and WHIM descriptors were the most influenced group, indicating that they are very sensitive to any conformational change in the molecular structure, i.e., high conformational dependence descriptors (Todeschini and Consonni, 2009).

Table 4.1. Number of descriptors affected ($r < 0.75$).

Method pairs	Descriptors				
	3D-MoRSE	WHIM	GETAWAY	Quantum chemical	Total
PM6 vs. HF	88	15	14	2	119
HF vs. DFT	35	13	11	0	59
PM6 vs. DFT	61	14	1	4	80

Different results were reported for the same descriptor classes for ionic liquids (Rybinska et al., 2016). Only one member of the GETAWAY descriptors (ISH; standardized information content on the leverage equality) was moderately different when PM6 and DFT were compared ($r = 0.60$). In case of the quantum chemical descriptors, energy-related parameters, such as hardness, exhibited significantly different values between PM6 and DFT as well as PM6 and HF. Interestingly, HF and DFT showed no significant effect, suggesting that thermodynamical parameters calculated at HF and DFT were closely related to each other (Table 4.1). For the PM6 vs. HF comparison, the most affected descriptors were from 3D-MoRSE group (Mor30s, Mor24i, Mor24u, Mor26s and Mor26i). Although to a lesser extent, a similar trend was observed for the PM6-DFT pair (Mor24i, Mor24u and Mor26s). Likewise, for HF vs. DFT, Mor32s, Mor30s, Mor26s and Mor27s seemed to be the most influenced descriptors (Table C.2).

Overall, the results demonstrated that geometry optimization method significantly affects the values of certain descriptors. A practical comparative reference summary for the sensitive/insensitive quantum chemicals and DRAGON descriptors was provided, for the first time (Önlü and Saçan, 2017a).

4.1.2. QSAR Model for Predicting HSA Binding Affinity: The Case Study

The log K_{HSA} values of a structurally heterogeneous dataset to HSA (Table A.1) were modeled as a case study considering the effect of different geometry optimization methods. While aiming at finding the best (robust, validated, predictive and the simplest) model, numerous MLR models with different training/test set divisions were developed using the DS1 obtained at the DFT level. Response-based division exhibited better results. Once the best model (M1_DFT) was established, the same model for PM6 (M1_PM6) and HF (M1_HF) was rebuilt using the DS1 generated at the PM6 and HF with the same setup, respectively. Since the best descriptors encoding HSA binding affinity were all insensitive to the geometry optimization method, identical models were obtained (Table 4.2). Consequently, the PM6-derived model was reported due to its lowest computational cost.

4-descriptor QSAR model for the prediction of HSA binding affinity together with the standard errors of coefficients based on the OLS estimates is presented in Equation 4.1.

$$\log K_{HSA} = -0.904\ (\pm 0.064) + 0.200\ (\pm 0.032)\ \text{nR10} - 0.291\ (\pm 0.057)\ \text{CATS2D\_01\_AN} + 0.320\ (\pm 0.059)\ \text{B10[C-N]} + 0.309\ (\pm 0.019)\ \text{ALOGP} \quad (4.1)$$

The model rigorously fulfils the up-to-date criteria regarding the fit, internal and external validation statistics (Table 4.2). In addition, based on the MAE_{test} parameter, given that the TSR = 2.73, calculated MAE_{test} = 0.14, and $MAE_{test} + 3*\sigma = 0.44$, the model further showed good prediction ability. A good accordance between predicted and experimental data is reflected by the homogenous distribution around the optimal line (Figure 4.2).

All descriptors but CATS2D_01_AN exhibited positive contribution to HSA binding. ALOGP is a measure of hydrophobicity/lipophilicity (Todeschini and Consonni, 2009). ALOGP with a positive coefficient accounts for increased HSA binding. Similarly, Chen and Chen (2012) reported ALOGP as the most important property encoding the binding affinity to HSA. A constitutional descriptor, nR10, accounts for the presence of either independent or fused 10-membered rings in molecules that are important features in determining physicochemical properties (Todeschini and Consonni, 2009). The positive sign of nR10 coefficient suggests that HSA binding increases with increasing number of 10-membered rings. A similar result was reported between the binding affinity of endocrine disrupting chemicals to human sex hormone binding globulin and nR10 (Liu et al., 2016). B10[C-N], determining the presence/absence of C-N at topological distance 10, exhibited a positive relationship to HSA binding. Similarly, B10[C-N] appeared as an important parameter describing the acute toxicity towards fathead minnow in another study (Wu et al., 2016). The last model descriptor CATS2D_01_AN is a 2D structure-based atom-pair descriptor encoding topological information where CATS denotes a chemically advanced template search. AN represents a hydrogen bond acceptor, negatively charged or ionizable as potential pharmacophore points of a molecular structure (Schneider et al., 1999). CATS2D_01_AN is inversely related to HSA binding. Predicted log K_{HSA} values, final training/test set status, as well as calculated model descriptor values are presented in Table A.1.

Table 4.2. Summary of case study models, statistical parameters, literature thresholds.

Fit and internal validation parameters[b, c]												
Model[a]	**Dataset**	**Descriptors**	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}	R^2_{Yscr}	Q^2_{Yscr}
M1_DFT	DS1	nR10 CATS2D_01_AN B10[C-N] ALOGP	0.828	0.818	0.237	0.906	83.964	0.794	0.259	0.888	0.054	-0.088
M1_HF	DS1	nR10 CATS2D_01_AN B10[C-N] ALOGP	0.828	0.818	0.237	0.906	83.964	0.794	0.259	0.888	0.053	-0.089
M1_PM6	**DS1**	**nR10 CATS2D_01_AN B10[C-N] ALOGP**	**0.828[d]**	**0.818**	**0.237**	**0.906**	**83.964**	**0.794[d]**	**0.259**	**0.888**	**0.053**	**-0.089**
M2_DFT	DS2	HOMT Mor29e HATS8p R5s	0.718	0.701	0.303	0.836	44.448	0.675	0.325	0.813	0.054	-0.089
M2_HF	DS2	HOMT Mor29e HATS8p R5s	0.665	0.646	0.330	0.799	34.704	0.615	0.354	0.770	0.055	-0.088
M2_PM6	DS2	HOMT Mor29e HATS8p R5s	0.657	0.637	0.334	0.793	33.503	0.602	0.360	0.761	0.054	-0.089
M3_DFT	DS3	VE1_H2 Mor24u R7s+ MLOGP2	0.763	0.749	0.278	0.865	56.263	0.733	0.295	0.849	0.054	-0.088
M3_HF	DS3	VE1_H2 Mor24u R7s+ MLOGP2	0.747	0.732	0.287	0.855	51.576	0.713	0.305	0.836	0.054	-0.089
M3_PM6	DS3	VE1_H2 Mor24u R7s+ MLOGP2	0.719	0.703	0.302	0.837	44.780	0.676	0.325	0.815	0.054	-0.091

External validation parameters[b, c]												
Model[a]	**Dataset**	R^2_{TEST}	$RMSE_{TEST}$	MAE_{TEST}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2_m	Δr^2_m	k	$(R^2 - R_0^2) / R^2$
M1_DFT	DS1	0.921	0.200	0.162	0.891	0.891	0.877	0.937	0.810	0.006	1.143	0.017
M1_HF	DS1	0.921	0.200	0.162	0.891	0.891	0.877	0.937	0.810	0.006	1.143	0.017
M1_PM6	**DS1**	**0.921[d]**	**0.200**	**0.162**	**0.891[d]**	**0.891[d]**	**0.877[d]**	**0.937[d]**	**0.810[d]**	**0.006[d]**	**1.143[d]**	**0.017[d]**
M2_DFT	DS2	0.760	0.306	0.274	0.746	0.745	0.713	0.839	0.737	0.023	1.149	0.003
M2_HF	DS2	0.806	0.303	0.275	0.750	0.750	0.718	0.834	0.713	0.013	1.288	0.019
M2_PM6	DS2	0.657	0.364	0.318	0.640	0.640	0.593	0.767	0.600	0.022	1.112	0.016
M3_DFT	DS3	0.849	0.279	0.227	0.788	0.788	0.761	0.855	0.840	0.004	1.362	0.000
M3_HF	DS3	0.780	0.334	0.241	0.696	0.696	0.657	0.775	0.752	0.028	1.481	0.001
M3_PM6	DS3	0.798	0.347	0.268	0.672	0.672	0.630	0.745	0.715	0.033	1.624	0.009

[a]DS1, all descriptors; DS2, only affected descriptors; DS3, DS1 except M1 descriptors; M1, proposed model in bold (Eq. 4.1); M2, comparison model of DS2; M3, comparison model of DS3; n_{TR}/n_{TEST} = 75/13; n_{TR}, number of training and n_{TEST}, number of test set compounds. [b]R^2, coefficient of determination; R^2_{adj}, adjusted R^2; $RMSE_{TR}$ and $RMSE_{CV}$, root mean squared error (*RMSE)* of training set and cross-validation *RMSE*; CCC_{TR} and CCC_{CV}, concordance correlation coefficient (*CCC*) of training set and cross-validation *CCC*; *F*, Fisher statistics; Q^2_{LOO}, leave-one-out cross-validation correlation coefficient; R^2_{Yscr} and Q^2_{Yscr}, new coefficients following Y-scrambling procedure; R^2_{TEST}, coefficient of determination of test set; $RMSE_{TEST}$, *RMSE* of test set; MAE_{TEST}: mean absolute error of test set; CCC_{TEST}, *CCC* of test set. [c]Literature thresholds and references are explained in Section 2.3.8. [d]Parameters passing the thresholds.

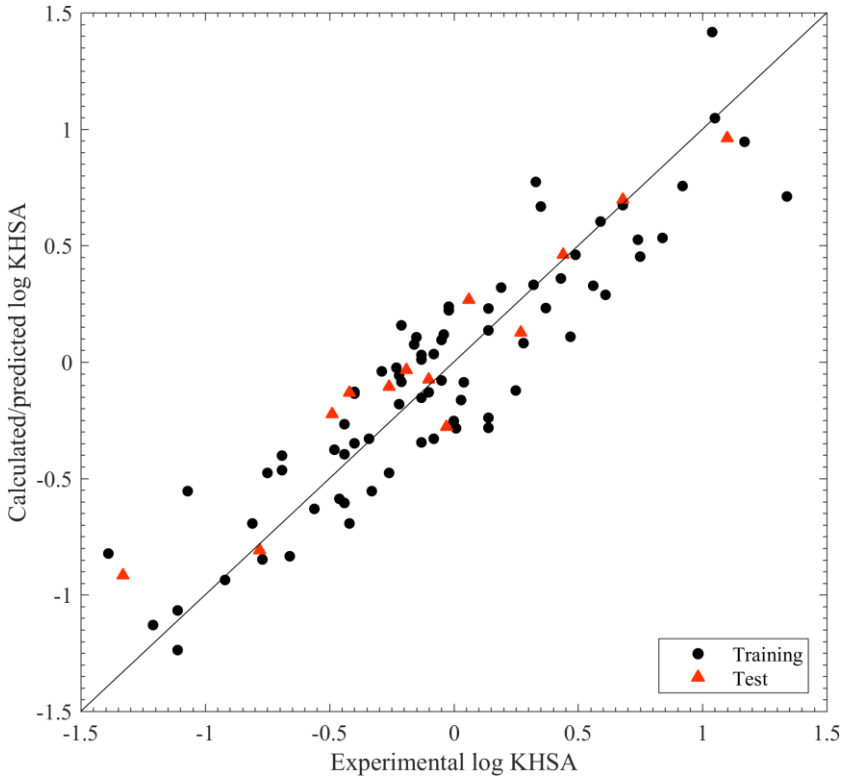

Figure 4.2. Predicted vs. experimental log K_{HSA} of the case study model.

The AD of the model was defined using the leverage and the standardization approaches (Figure 4.3). Remarkably, no response outlier for the training and the test set compounds was identified, suggesting that the predictions of HSA binding affinity were reliably interpolated by the proposed model. It is noteworthy that only two compounds (cromolyn and methotrexate) exhibited slightly higher leverage values than the critical one, thus, were identified as *"good leverage"* compounds reinforcing the model due to correct extrapolation (Gramatica, 2007). Cromolyn and methotrexate are the only compounds of the dataset having the maximum value of 2 for CATS2D_01_AN. In addition, cromolyn appeared to be the only chemical highlighted by the standardization approach (Figure 4.3, indicated by an oval boundary).

Finally, the model was applied to an external set of 89 pharmaceutically active compounds with no experimental log K_{HSA} selected from the literature (Table A.2). The model provided remarkable prediction coverage of 97% (Figure 4.4). Notably, only one compound (amidosulfonic acid) was found to be structurally different, and its prediction was extrapolated. Likewise, predicted log K_{HSA} of two other compounds (acyclovir and diethanolamine) were slightly outside the model prediction range of −1.24 to 1.42.

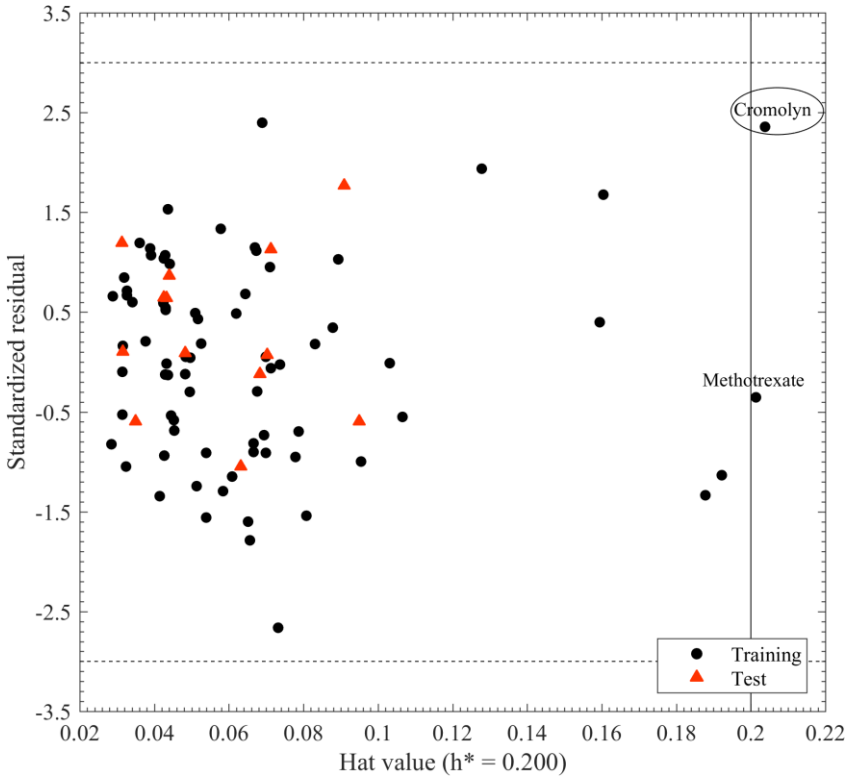

Figure 4.3. Applicability domain of the case study model.

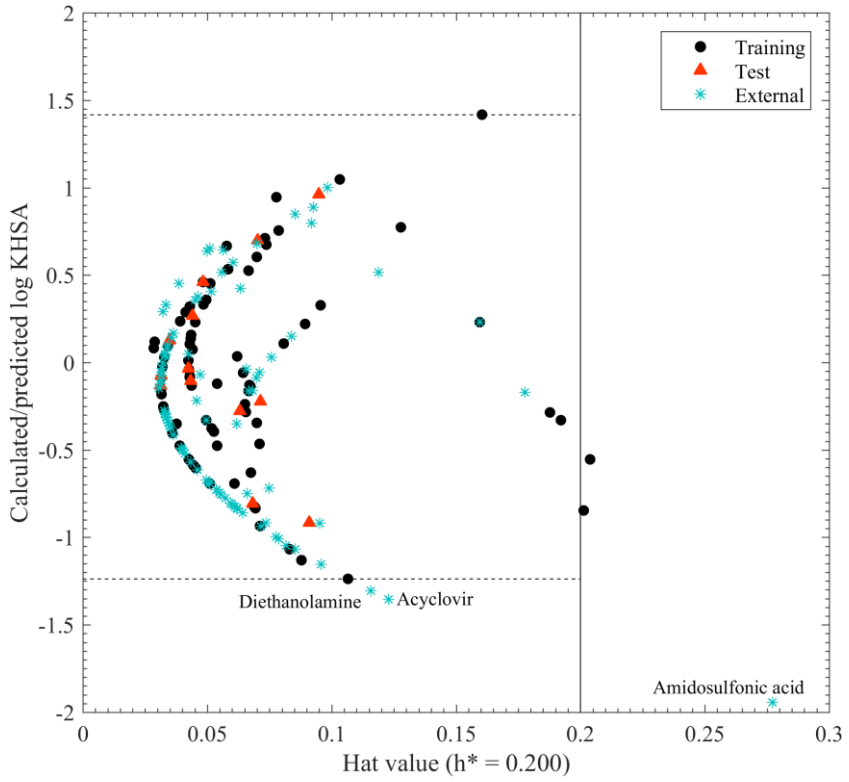

Figure 4.4. Prediction coverage of the case study model.

Calculated ALOGP of all three compounds were lower than its range. Model descriptor ranges are as follows: nR10: 0 to 3, CATS2D_01_AN: 0 to 2, B10[C-N]: 0 to 1, and ALOGP: –1.111 to 6.467. A comprehensive summary of earlier efforts towards QSAR modelling for binding affinity to HSA is available online as SM at the article's web page http://dx.doi.org/10.1080/1062936X.2017.1343253 (Önlü and Saçan, 2017a).

4.1.3. Geometry Optimization Method versus Model Performance

Two datasets, DS2 and DS3, were used to develop two sets of 4-descriptor MLR models labelled as M2 and M3, respectively. The same training/test set division of the proposed model was employed (Table A.1). The results enabled to examine the comparative influence of using descriptors obtained from different quantum chemical methods (PM6, HF and DFT) on the statistical quality and prediction power of QSAR models (Table 4.2). M2 models were generated with descriptors having low sensitivity ($0.75 \leq r < 1$) such as HOMT, Mor29e, HATS8p and R5s. Regarding the fit and internal validation metrics, DFT-based model yielded better results. AD of each model was defined individually. With the highest calculated ALOGP, itraconazole appeared as the only common influential compound in all M2 models. M3 models provided a broader comparison possibility for the different sensitiveness of the descriptors towards the quantum chemical methods. Two insensitive (VE1_H2 and MLOGP2), one sensitive (R7s+) and a highly sensitive (Mor24u) descriptors appeared in M3 models. Going from PM6 to higher levels, the inclusion of a highly sensitive descriptor, Mor24u, resulted in a meaningful improvement in the value of R^2 that is greater than the breakpoint criterion of 0.02. Concerning overall statistical quality and model performance, the DFT-based model yielded better results. AD of each M3 model was also defined. Predicted log K_{HSA} by M2 and M3 models, calculated model descriptors at each level together with the coefficients, as well as the graphs for all models are available online as SM (http://dx.doi.org/10.1080/1062936X.2017.1343253).

The results revealed that there is no golden standard of the required optimization method for a better definition of geometry. Likewise, no particular quantum chemical level of the theory is universally superior to another. For the models built with descriptors not affected by the geometry optimization method, models based on PM6-derived descriptors provide satisfactory results. In case of the models with sensitive descriptors, although the quality of predictions generally increases with the improvement in the accuracy of the geometry optimization method, the essential determining factor is

the significance level of the influence on that particular descriptor. In order to assess the impact of the geometry optimization method on established QSAR models, the QSAR DataBank (QsarDB) repository (Ruusmann et al., 2015) was scanned for the models built with DRAGON descriptors. At the time of writing, QsarDB provided 37 models collected from the literature for various endpoints, spanning from different physicochemical properties to toxicity as well as binding affinity to oestrogen receptor (Table 4.3). Noticeably, 16 of the published models (43%) included descriptors affected by the geometry optimization method, further emphasizing the importance of the approach.

The curiosity for finding the most accurate geometry could be bridged by the notion from the book Zen and the Art of Motorcycle Maintenance: An Inquiry into Values by Robert M. Pirsig (1974): *"One geometry cannot be more true than another; it can only be more convenient. Geometry is not true, it is advantageous"*. When all this is taken into account, it might be more appropriate not to recommend a certain level of the theory. Therefore, a number of steps through a rational approach on the selection of geometry optimization method for QSAR model development were presented for the first time (Figure 4.5).

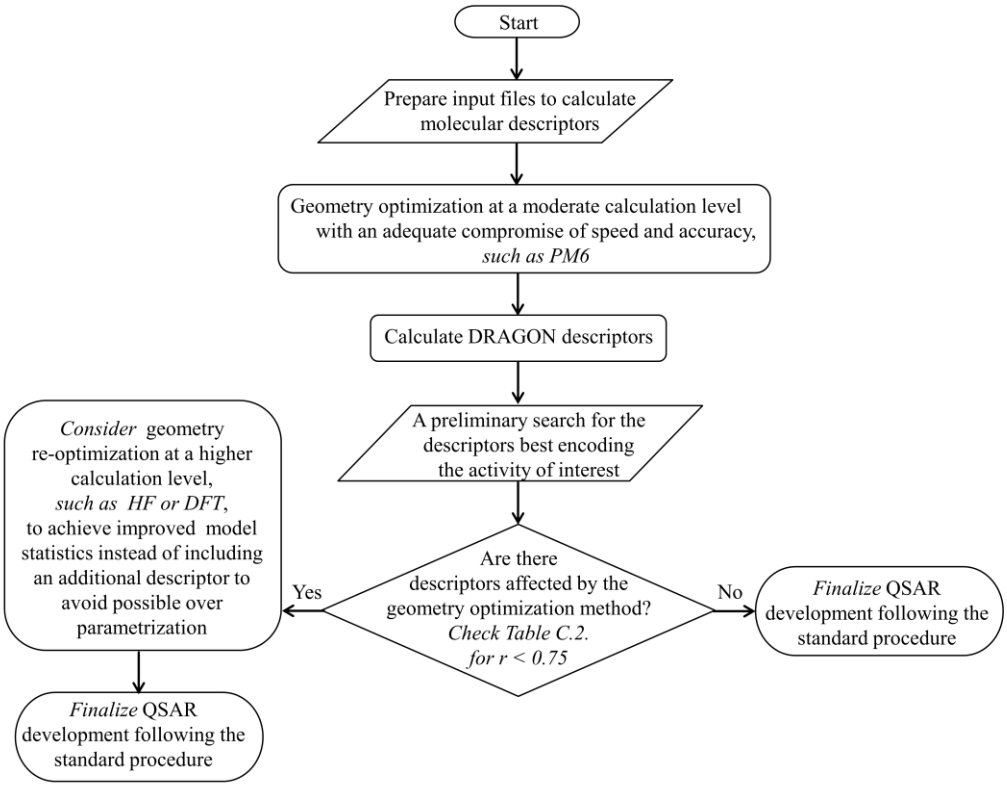

Figure 4.5. Rational approach for QSAR development.

Table 4.3. Different QSAR models using DRAGON descriptors.

Model	Activity/Property	Model descriptors	Reference
1	Flash point	IVDM, G2e, nRNH2, Hy	Khajeh and Modarress, 2011
2	Relative binding affinity to estrogen receptor α (ERα)	E1s, MATS1v, L3s, Mor12v, RDF020e	Li et al., 2012
3	Nucleoside activity against Leishmania donovani	Mor26v, Gap	Oliveira and Takahata, 2008
4	Aqueous solubility as logWS [mg/L]	AMW, CIC0, MATS7e	Bhhatarai and Gramatica, 2011a
5	Vapour pressure as logVP [mm Hg]	B09[N-Cl], RBN, BELp2/SpMin2_Bh(p)	Bhhatarai and Gramatica, 2011a
6	Melting point [°C]	X1A, GGI4, R2e, F03[N-N]	Bhhatarai and Gramatica, 2011a
7	Octanol/water partition coefficient as $\log K_{ow}$	nN, MATS1v, GATS3m, B08[C-C]	Bhhatarai and Gramatica, 2011a
8	Mouse oral toxicity as $\log(LD_{50})$ [-log(mmol/kg)]	HATS2u, B09[C-O], F01[C-O], B04[C-F]	Bhhatarai and Gramatica, 2010a
9	Rat oral toxicity as $\log(1/LD_{50})$ [-log(mmol/kg)]	D/Dtr09, MATS1e, E1u, H8m	Bhhatarai and Gramatica, 2010a
10	Mice acute inhalation toxicity as pLC_{50} [-log(mmol/m^3)]	MLOGP, X3v, H-048, F01[C-C]	Bhhatarai and Gramatica, 2010b
11	Rat acute inhalation toxicity as pLC_{50} [-log(mmol/m^3)]	Jhetv/J_Dz(e), PCR, MLOGP, B02[Cl-Cl]	Bhhatarai and Gramatica, 2010b
12	Vapour pressure as logVP [log(mm Hg)]	F03[C-F], nDB, AAC	Bhhatarai and Gramatica, 2011b
13	Critical micelle concentration as logCMC [log(mol/L)]	X3	Bhhatarai and Gramatica, 2011b
14	Aqueous solubility as logAqS [log(mg/L)]	T(F..F), SIC1	Bhhatarai and Gramatica, 2011b
15	Fish bioconcentration factor as logBCF	IDDM, HIC, nHAcc, GATS1e, MATS1p	Gramatica and Papa, 2005
16	96-h Fathead minnow toxicity as $\log(1/LC_{50})$ [log(L/mmol)]	WA /WiA_D, Mv, H-046, nCb-, MAXDP, nN	Papa et al., 2005
17	Logarithm of soil sorption coefficient $\log K_{oc}$	VED1/VE1_D, nHAcc, MAXDP, CIC0	Gramatica et al., 2007a
18	Mutagenicity potency in TA100 without the S9 activation system [log(revertants/nmol)]	CIC1, PW2	Gramatica et al., 2007b
19	Mutagenicity on human h1A1v2 cells	S1K, nArNO2	Papa et al., 2008
20	Normal boiling point [K]	R1e+, MATS1m, X1sol, Me, ESpm02d/SM02_AEA(dm)	Abooali and Sobati, 2014
21	Enthalphy of vaporization at normal boiling point [kJ/kg]	Hy, P1s, ATS1m, ALOGPS_logP, ZM1V, Har/Wi_H2	Abooali and Sobati, 2014
22	Global half-life index	X0v, Mv, MAXDP, nHDon, CIC0, O-060	Gramatica and Papa, 2007
23	Estrogen receptor relative binding affinity	X2A, TIC1, EEig02d/, Eig02_AEA(dm), JGI10, SPH, E1u, RTm+, nArOR	Liu et al., 2006
24	Self-accelerating decomposition temperature [°C]	MAXDP, RDF095v, R2m+, nHDon, C-004, C-006	Pan et al., 2014
25	Degradation by NO3 radicals as -logkNO3 [log(s molecule/cm3)]	HOMO, nBnz, Me	Papa and Gramatica, 2008
26	Henry's law constant as logH [Pa.m3/mol]	BEHe7/SpMax7_Bh(e)	Papa et al., 2009
27	Melting point [°C]	X2A	Papa et al., 2009
28	Subcooled liquid vapour pressure as $\log P_L$ [Pa]	T(O..Br)	Papa et al., 2009
29	Water solubility as $\log S_w$ [mol/L]	Mor23m	Papa et al., 2009
30	Octanol-air partition coefficient as logKoa	T(O..Br)	Papa et al., 2009
31	Octanol-water partition coefficient as logKow	T(O..Br)	Papa et al., 2009
32	Flash point [K]	Ss, Jhetv/J_Dz(e), HATS1m	Bagheri et al., 2012

Table 4.3. Continued.

Model	Activity/Property	Model descriptors	Reference
33	48-h Honeybee toxicity as -log(LD_{50}) [log(bee/micromol)]	Dipole_X_DMol3, GATS6e, CIC2, SIC2, Shadow_XZfrac, Shadow_YZfrac, LUMO_Energy_DMol3, C-040, B01[S-P], B05[C-C], GATS1m, MATS1m, TIE, AAC, BELv3/SpMin3_Bh(v), Inflammat-50	Dulin et al., 2012
34	72-h Algal toxicity as log(1/EC_{50}) [-log(mol/L)]	AEigZ/SpAbs_Dz(Z), T(N..S), SEigV/SM1_Dz(v)	Gramatica et al., 2012
35	Persistency, bioaccumulation and toxicity (PBT) Index	nX, nBM, nHDon, MAXDP	Papa and Gramatica, 2010
36	Degradation by OH radicals as -logk(OH) [log(s/cm3)]	HOMO, nX, nCbH, IDE	Roy et al., 2011
37	Rodent carciogenicity [log(mmol/kg/d)]	nRNNOx, MAXDP, Cl-089, Mp, nCXr, nArOR, nArX, C-005	Kar et al., 2012

4.2. QSTR Models for Toxicity Prediction of Algae

4.2.1. Linear Models

First, two MoA-based local MLR models were constructed for Group 1 (61 non-polar narcotics) and Group 2 (74 polar narcotics). Local model 1 (LM1) for non-polar narcotics and local model 2 (LM2) for polar narcotics were given together with the standard errors of coefficients based on the OLS estimates in Equations 4.2 and 4.3, respectively.

$$pEC_{50} = 1.233\ (\pm 0.216) - 2.234\ (\pm 0.445)\ SIC0 + 1.014\ (\pm 0.183)\ R2v + 0.991\ (\pm 0.045)\ MLOGP \quad (4.2)$$

$$pEC_{50} = -2.295\ (\pm 0.415) + 0.453\ (\pm 0.113)\ MAXDN + 26.170\ (\pm 2.122)\ X2Av + 1.200\ (\pm 0.211)\ Mor11m + 0.194\ (\pm 0.061)\ CATS2D\_05\_DL \quad (4.3)$$

Chemicals, experimental and predicted pEC_{50} values, final training/test set status, as well as calculated model descriptor values are presented in Table 4.4 and 4.5 for Group 1 and Group 2, respectively. The local models are decent regarding the fit, internal, and external validation statistics (Table 4.6). Descriptor types and chemical meanings are provided in Table 4.7. Good agreements between predicted and experimental data are visible from the homogenous distribution around the optimal line for both local models (Figure 4.6 (A) and 4.7 (A)). The ADs of the models were defined using the leverage and the standardization approaches (Figure 4.6 (B) and 4.7 (B)). Chemicals highlighted by the standardization approach were indicated by an oval boundary. For LM1, carbon tetrachloride appeared as an outlier, having the highest value for R2v. Dichloromethane, having the highest SIC0 value was identified as an outlier by the standardization approach but not by the leverage approach. For LM2, 4-n-octylphenol, with the highest value for X2Av and Mor11m was identified as an outlier.

The local models are promising not only for being comparable to the literature (Table 4.8) but also for providing rigorous validation metrics. Additionally, Fu et al. (2015) searched for the relationship of toxicity to the log K_{ow} for non-polar narcotics. Although this MoA-based toxicity relationship is useful, it is limited to the number of chemicals in that group.

Table 4.4. Group 1 chemicals, experimental/predicted pEC_{50}, model descriptors.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (LM1)	SIC0	R2v	MLOGP	HPV Status[a]
1	110-82-7	Cyclohexane	3.64	4.42	0.22	0.577	3.124	HPV
2	108-87-2	Methylcyclohexane	5.46	4.77	0.21	0.552	3.477	HPV
3	1678-91-7	Ethylcyclohexane	5.25	5.10	0.20	0.527	3.811	HPV
5	109-69-3	1-chlorobutane	3.41	3.14	0.32	0.396	2.226	HPV
6	75-09-2	Dichloromethane	2.04	1.84	0.66	0.710	1.364	HPV
7	75-34-3	1,1-dichloroethane*	2.97	2.60	0.50	0.678	1.817	HPV
8	107-06-2	1,2-dichloroethane	2.57	2.42	0.50	0.492	1.817	HPV
9	78-87-5	1,2-dichloropropane	2.94	3.08	0.42	0.556	2.226	HPV
10	142-28-9	1,3-dichloropropane*	3.04	3.04	0.42	0.523	2.226	N
12	67-66-3	Trichloromethane	2.71	2.95	0.59	1.219	1.817	HPV
14	79-00-5	1,1,2-trichloroethane	3.11	3.04	0.52	0.754	2.226	HPV
15	96-18-4	1,2,3-trichloropropane	3.27	3.45	0.45	0.620	2.604	HPV
16	56-23-5	Carbon tetrachloride	4.84	4.65	0.31	1.878	2.226	HPV
18	79-34-5	1,1,2,2-tetrachloroethane	3.49	3.66	0.50	0.946	2.604	HPV
19	76-01-7	Pentachloroethane	4.38	4.52	0.43	1.302	2.957	HPV
43	71-36-3	1-butanol	1.68	1.80	0.30	0.430	0.800	HPV
44	78-83-1	Iso-Butanol	1.64	1.75	0.30	0.381	0.800	HPV
46	75-65-0	2-methyl-2-propanol	1.66	1.75	0.30	0.378	0.800	HPV
47	71-41-0	1-pentanol*	2.38	2.27	0.27	0.440	1.209	HPV
48	584-02-1	3-pentanol	2.13	2.24	0.27	0.406	1.209	N
49	111-27-3	Hexanol	2.95	2.69	0.25	0.445	1.587	HPV
50	111-70-6	1-heptanol*	3.53	3.08	0.24	0.456	1.940	N
51	111-87-5	1-octanol	3.67	4.35	0.23	0.466	3.186	HPV
52	143-08-8	1-nonanol	4.82	4.70	0.22	0.475	3.503	HPV
53	112-30-1	1-decanol	5.16	5.02	0.21	0.479	3.806	HPV
54	25339-17-7	Isodecyl alcohol	4.37	4.23	0.21	0.591	2.894	HPV
56	108-93-0	Cyclohexanol	2.39	2.40	0.28	0.585	1.195	HPV
99	60-29-7	Diethylether*	1.51	1.80	0.30	0.429	0.800	HPV
100	142-96-1	1,1'-oxybis-butane	3.77	3.48	0.23	0.492	2.274	HPV
101	111-44-4	Bis(2-chloroethyl) ether*	2.62	2.45	0.42	0.574	1.587	HPV
102	127-90-2	2,3,3,3,2',3',3',3'-Octachlorodipropyl ether*	5.50	5.55	0.40	1.240	4.002	N
114	67-64-1	Acetone	0.96	0.92	0.39	0.354	0.202	HPV
119	693-54-9	2-decanone	4.50	4.81	0.22	0.435	3.661	N
120	112-12-9	2-undecanone*	4.95	5.12	0.21	0.443	3.953	N
121	593-08-8	2-tridecanone	6.22	5.72	0.20	0.457	4.506	N
294	71-43-2	Benzene	3.12	3.26	0.28	0.407	2.255	HPV
295	108-88-3	Toluene	3.53	3.67	0.26	0.418	2.608	HPV

Table 4.4. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (LM1)	SIC0	R2v	MLOGP	HPV Status[a]
296	100-41-4	Ethylbenzene	4.36	4.17	0.24	0.542	2.942	HPV
298	95-47-6	o-Xylene	4.37	4.03	0.24	0.404	2.942	HPV
299	108-38-3	m-Xylene*	4.08	4.03	0.24	0.404	2.942	HPV
300	106-42-3	p-Xylene	3.83	4.05	0.24	0.427	2.942	HPV
301	103-65-1	n-Propylbenzene	4.82	4.55	0.22	0.576	3.259	HPV
302	98-82-8	Isopropylbenzene	4.66	4.46	0.22	0.490	3.259	HPV
303	104-51-8	Butylbenzene	4.92	4.87	0.21	0.576	3.562	N
304	99-87-6	p-Cymene	4.36	4.78	0.21	0.487	3.562	HPV
305	98-51-1	4-tert-Butyltoluene	4.53	5.08	0.21	0.474	3.854	HPV
306	25321-09-9	Diisopropylbenzene	4.68	5.40	0.20	0.505	4.135	HPV
308	827-52-1	Cyclohexylbenzene	5.37	5.16	0.21	0.666	3.743	N
309	108-90-7	Chlorobenzene	3.58	3.74	0.37	0.479	2.876	HPV
310	95-49-8	2-chlorotoluene	4.21	4.15	0.33	0.459	3.210	HPV
311	108-41-8	3-chlorotoluene	4.55	4.17	0.33	0.485	3.210	N
312	106-43-4	4-chlorotoluene	4.45	4.18	0.33	0.494	3.210	HPV
314	95-50-1	1,2-dichlorobenzene	4.01	4.32	0.41	0.539	3.478	HPV
315	541-73-1	1,3-dichlorobenzene	4.58	4.32	0.41	0.545	3.478	HPV
316	106-46-7	1,4-dichlorobenzene	4.19	4.33	0.41	0.548	3.478	HPV
317	95-73-8	2,4-dichlorotoluene	4.80	4.72	0.37	0.533	3.795	HPV
318	95-75-0	3,4-dichlorotoluene	4.93	4.74	0.37	0.553	3.795	N
319	19398-61-9	2,5-dDichlorotoluene	4.98	4.71	0.37	0.524	3.795	N
320	118-69-4	2,6-dichlorotoluene	4.78	4.68	0.37	0.500	3.795	N
321	87-61-6	1,2,3-trichlorobenzene	5.05	4.93	0.42	0.596	4.063	N
323	120-82-1	1,2,4-trichlorobenzene	4.57	4.94	0.42	0.604	4.063	HPV

*Test set compound. [a]Production volume status according to OECD (2009). HPV: High production volume. N: Not HPV.

Table 4.5. Group 2 chemicals, experimental/predicted pEC_{50}, model descriptors.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (LM2)	MAXDN	X2Av	Mor11m	CATS2D_05_DL	HPV Status[a]
404	108-95-2	phenol	2.73	2.91	1.35	0.167	0.186	0	HPV
407	106-44-5	4-cresol	3.30	3.57	1.34	0.184	0.206	1	HPV
408	108-39-4	3-methylphenol*	2.87	3.34	1.33	0.184	0.186	0	HPV
410	95-48-7	2-methylphenol	2.93	3.22	1.30	0.179	0.199	0	HPV
411	90-00-6	2-ethylphenol	3.59	3.33	1.26	0.180	0.286	0	HPV
412	123-07-9	4-ethylphenol	3.75	3.72	1.33	0.184	0.341	1	HPV
413	620-17-7	3-ethylphenol*	3.48	3.68	1.31	0.184	0.310	1	HPV
414	526-75-0	2,3-dimethylphenol*	3.40	3.40	1.28	0.185	0.227	0	HPV
415	576-26-1	2,6-dimethylphenol	3.44	3.53	1.25	0.187	0.305	0	HPV
416	95-65-8	3,4-dimethylphenol	3.58	3.65	1.32	0.189	0.172	1	HPV
417	95-87-4	2,5-dimethylphenol*	3.60	3.63	1.28	0.191	0.290	0	HPV
418	105-67-9	2,4-dimethylphenol	3.96	3.82	1.29	0.191	0.283	1	HPV
419	108-68-9	3,5-dimethylphenol	3.65	3.61	1.31	0.195	0.175	0	HPV
422	527-60-6	2,4,6-trimethylphenol	4.15	3.99	1.25	0.196	0.329	1	HPV
423	697-82-5	2,3,5-trimethylphenol*	4.00	3.61	1.27	0.195	0.194	0	N
424	2416-94-6	2,3,6-trimethylphenol	3.98	3.61	1.24	0.191	0.286	0	HPV
425	88-18-6	2-tert-butyl phenol	5.06	4.60	1.28	0.235	0.138	0	HPV
428	89-72-5	o-sec-butylphenol	4.34	4.20	1.25	0.205	0.306	1	HPV
429	89-83-8	thymol	4.03	4.21	1.25	0.214	0.279	0	HPV
430	99-71-8	p-sec-butylphenol	4.30	4.21	1.32	0.207	0.243	1	HPV
431	14938-35-3	4-pentylphenol*	4.87	5.06	1.31	0.223	0.610	1	N
432	88-60-8	6-tert-butyl-m-cresol	4.94	4.71	1.26	0.237	0.197	0	HPV
433	2219-82-1	6-tert-butyl-o-cresol	4.42	4.72	1.23	0.235	0.257	0	HPV
434	2409-55-4	2-tert-butyl-p-cresol	4.96	4.87	1.27	0.237	0.161	1	HPV
435	1879-09-0	2-(1,1-dimethylethyl)-4,6-dimethylphenol	4.40	4.97	1.24	0.236	0.282	1	N
436	96-76-4	2,4-di-tert-butylphenol*	5.31	5.27	1.27	0.260	-0.007	1	HPV
437	1806-26-4	4-n-octylphenol	6.17	6.06	1.30	0.246	0.947	1	N
438	5510-99-6	2,6-di-sec-butylphenol	5.27	4.80	1.15	0.220	0.355	2	N
441	120-95-6	2,4-di-tert-pentylphenol	5.14	5.25	1.24	0.239	0.283	2	HPV
443	95-57-8	2-chlorophenol	3.39	3.58	1.53	0.186	0.259	0	HPV
444	108-43-0	3-chlorophenol	4.05	3.94	1.46	0.192	0.454	0	N
445	106-48-9	4-chlorophenol	3.86	4.08	1.42	0.191	0.448	1	HPV
446	59-50-7	4-chloro-3-methylphenol	3.98	4.06	1.41	0.195	0.354	1	HPV
447	576-24-9	2,3-dichlorophenol	4.17	4.16	1.65	0.197	0.465	0	N
448	120-83-2	2,4-dichlorophenol	4.42	4.53	1.61	0.203	0.490	1	HPV
449	583-78-8	2,5-dichlorophenol	4.65	4.20	1.65	0.203	0.361	0	HPV
450	87-65-0	2,6-dichlorophenol	4.01	4.17	1.72	0.199	0.397	0	N
451	95-77-2	3,4-dichlorophenol	4.87	4.60	1.54	0.201	0.624	1	N
452	591-35-5	3,5-dichlorophenol	4.89	4.63	1.58	0.208	0.644	0	N
453	15950-66-0	2,3,4-trichlorophenol	4.68	4.74	1.73	0.204	0.599	1	N
454	933-78-8	2,3,5-trichlorophenol*	4.94	4.65	1.77	0.210	0.546	0	N
455	933-75-5	2,3,6-trichlorophenol	4.39	4.53	1.84	0.206	0.500	0	N
456	95-95-4	2,4,5-trichlorophenol	4.77	4.75	1.73	0.210	0.481	1	N
458	88-06-2	2,4,6-trichlorophenol	4.54	4.99	1.80	0.212	0.606	1	HPV
508	62-53-3	aniline	3.09	2.99	0.85	0.176	0.246	0	HPV
509	95-53-4	2-methylaniline	2.97	3.26	0.80	0.186	0.268	0	HPV
510	108-44-1	3-methylaniline	3.60	3.38	0.83	0.191	0.253	0	HPV
511	106-49-0	4-methylaniline	3.53	3.63	0.84	0.191	0.294	1	HPV
512	578-54-1	2-ethylaniline	3.39	3.38	0.76	0.187	0.361	0	N
513	587-02-0	3-ethylaniline	3.93	3.73	0.81	0.191	0.387	1	N
514	589-16-2	4-ethylaniline	4.14	3.77	0.83	0.190	0.441	1	N
515	87-59-2	2,3-dimethylaniline	3.53	3.39	0.78	0.191	0.275	0	N

Table 4.5. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (LM2)	MAXDN	X2Av	Mor11m	CATS2D_05_DL	HPV Status[a]
516	87-62-7	2,6-dimethylaniline	3.05	3.43	0.75	0.192	0.301	0	HPV
517	95-64-7	3,4-dimethylaniline	4.19	3.70	0.82	0.195	0.269	1	N
518	95-68-1	2,4-dimethylaniline	3.49	3.79	0.79	0.197	0.317	1	HPV
519	95-78-3	2,5-dimethylaniline	3.42	3.58	0.78	0.197	0.304	0	N
520	108-69-0	3,5-dimethylaniline	3.63	3.66	0.81	0.202	0.251	0	N
521	579-66-8	2,6-diethylaniline	3.56	3.69	0.68	0.193	0.525	0	HPV
522	99-88-7	4-isopropylaniline	3.88	4.32	0.83	0.217	0.310	1	HPV
523	88-05-1	2,4,6-trimethylaniline	3.76	3.91	0.75	0.201	0.342	1	N
525	95-51-2	2-chloroaniline*	3.59	3.64	1.05	0.193	0.344	0	HPV
526	108-42-9	3-chloroaniline	3.76	3.96	0.98	0.199	0.502	0	N
527	106-47-8	4-chloroaniline*	4.55	4.12	0.95	0.199	0.486	1	N
528	95-81-8	2-chloro-5-methylaniline	4.46	3.90	1.04	0.203	0.342	0	N
529	95-76-1	3,4-dichloroaniline	4.61	4.63	1.17	0.207	0.658	1	HPV
530	95-82-9	2,5-dichloroaniline	4.11	4.27	1.15	0.209	0.483	0	N
531	554-00-7	2,4-dichloroaniline*	4.61	4.57	1.16	0.209	0.566	1	N
532	608-27-5	2,3-dichloroaniline	4.38	4.07	1.24	0.202	0.432	0	N
533	608-31-1	2,6-dichloroaniline	3.84	4.01	1.22	0.204	0.345	0	N
534	626-43-7	3,5-dichloroaniline	4.57	4.70	1.10	0.215	0.730	0	N
535	634-67-3	2,3,4-trichloroaniline	4.74	4.66	1.36	0.209	0.561	1	N
536	634-93-5	2,4,6-trichloroaniline	4.67	4.84	1.30	0.217	0.559	1	N
537	636-30-6	2,4,5-trichloroaniline	4.80	4.83	1.25	0.215	0.618	1	N
538	634-91-3	3,4,5-trichloroaniline	5.14	5.03	1.32	0.214	0.775	1	N

*Test set compound. [a]Production volume status according to OECD (2009). HPV: High production volume. N: Not HPV.

Table 4.6. Summary of algae models, statistical parameters, literature thresholds.

Fit and internal validation parameters[b, c]												
Model[a]	**Number of descriptors**	n_{TR}/n_{TEST}	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}	R^2_{Yscr}	Q^2_{Yscr}
LM1	3	52/9	0.931[d]	0.927	0.298	0.964	216.126	0.919[d]	0.322	0.958	0.058	-0.128
LM2	4	63/11	0.861[d]	0.851	0.246	0.925	89.690	0.834[d]	0.269	0.910	0.064	-0.107
GM	8	389/66	0.661[d]	0.654	0.653	0.796	92.682	0.643[d]	0.670	0.786	0.021	-0.027
CPANN	8	389/66	0.900[d]	–	–	–	–	0.560[d]	–	–	–	–

External validation parameters[b, c]												
	R^2_{TEST}	$RMSE_{TEST}$	MAE_{TEST} (95% data)	$MAE_{TEST} + 3*\sigma$ (95% data)	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2_m	Δr^2_m	k	$(R^2 - R_0^2) / R^2$
LM1	0.967[d]	0.234	–	–	0.968[d]	0.962[d]	0.957[d]	0.981[d]	0.922[d]	0.032[d]	1.013[d]	0.004[d]
LM2	0.887[d]	0.257	0.166[d]	0.641[d]	0.882[d]	0.882[d]	0.848[d]	0.933[d]	0.777[d]	0.116[d]	1.007[d]	0.004[d]
GM	0.718[d]	0.607	0.438[d]	1.266[d]	0.715[d]	0.713[d]	0.707[d]	0.837	0.606[d]	0.171[d]	1.015[d]	0.002[d]
CPANN	0.650[d]	–	–	–	–	–	–	–	–	–	–	–

[a]LM1: Local Model 1 (Eq. 4.2); LM2: Local Model 2 (Eq. 4.3); GM: Global Model (Eq. 4.4); CPANN: Counter propagation artificial neural network model; n_{TR}, number of training and n_{TEST}, number of test set compounds. [b]R^2, coefficient of determination; R^2_{adj}, adjusted R^2; $RMSE_{TR}$ and $RMSE_{CV}$, root mean squared error (*RMSE)* of training set and cross-validation *RMSE*; CCC_{TR} and CCC_{CV}, concordance correlation coefficient (*CCC*) of training set and cross-validation *CCC*; *F*, Fisher statistics; Q^2_{LOO}, leave-one-out cross-validation correlation coefficient; R^2_{Yscr} and Q^2_{Yscr}, new coefficients following Y-scrambling procedure; R^2_{TEST}, coefficient of determination of test set; $RMSE_{TEST}$, *RMSE* of test set; MAE_{TEST}: mean absolute error for 95% of test set data; σ: standard deviation of the absolute error for test set data; CCC_{TEST}, *CCC* of test set. [c]Literature thresholds and references are explained in Section 2.3.8. [d]Parameters passing the thresholds.

Table 4.7. Descriptors of algae models, types and chemical meanings.

Descriptor	Type	Chemical meaning	Model[a]
SIC0	Information indices	Structural Information Content index (neighborhood symmetry of 0-order)	LM1
R2v	GETAWAY descriptors	R autocorrelation of lag 2/weighted by van der Waals volume	LM1
MLOGP	Molecular properties	Moriguchi octanol-water partition coefficient (log P)	LM1
MAXDN	Topological indices	Maximal electrotopological negative variation	LM2
X2Av	Connectivity indices	Average valence connectivity index of order 2	LM2
Mor11m	3D-MoRSE descriptors	Signal 11/weighted by mass	LM2
CATS2D_05_DL	CATS 2D	CATS2D Donor-Lipophilic at lag 05	LM2
SPAM	Geometrical descriptors	Average span R	GM/CPANN
Mor31p	3D-MoRSE descriptors	Signal 31/weighted by polarizability	GM/CPANN
NdsCH	Atom-type E-state indices	Number of atoms of type dsCH	GM/CPANN
CATS2D_02_AP	CATS 2D	CATS2D Acceptor-Positive at lag 02	GM/CPANN
B05[C-S]	2D Atom Pairs	Presence/absence of C-S at topological distance 5	GM/CPANN
F03[C-N]	2D Atom Pairs	Frequency of C-N at topological distance 3	GM/CPANN
MLOGP2	Molecular properties	Squared Moriguchi octanol-water partition coefficient	GM/CPANN
Hardness	Quantum chemical (energy)	Half of the energy difference between the lowest unoccupied and highest occupied molecular orbitals	GM/CPANN

[a]LM1: Local Model 1 (Eq. 4.2); LM2: Local Model 2 (Eq. 4.3); GM: Global Model (Eq. 4.4); CPANN: Counter propagation artificial neural network model.

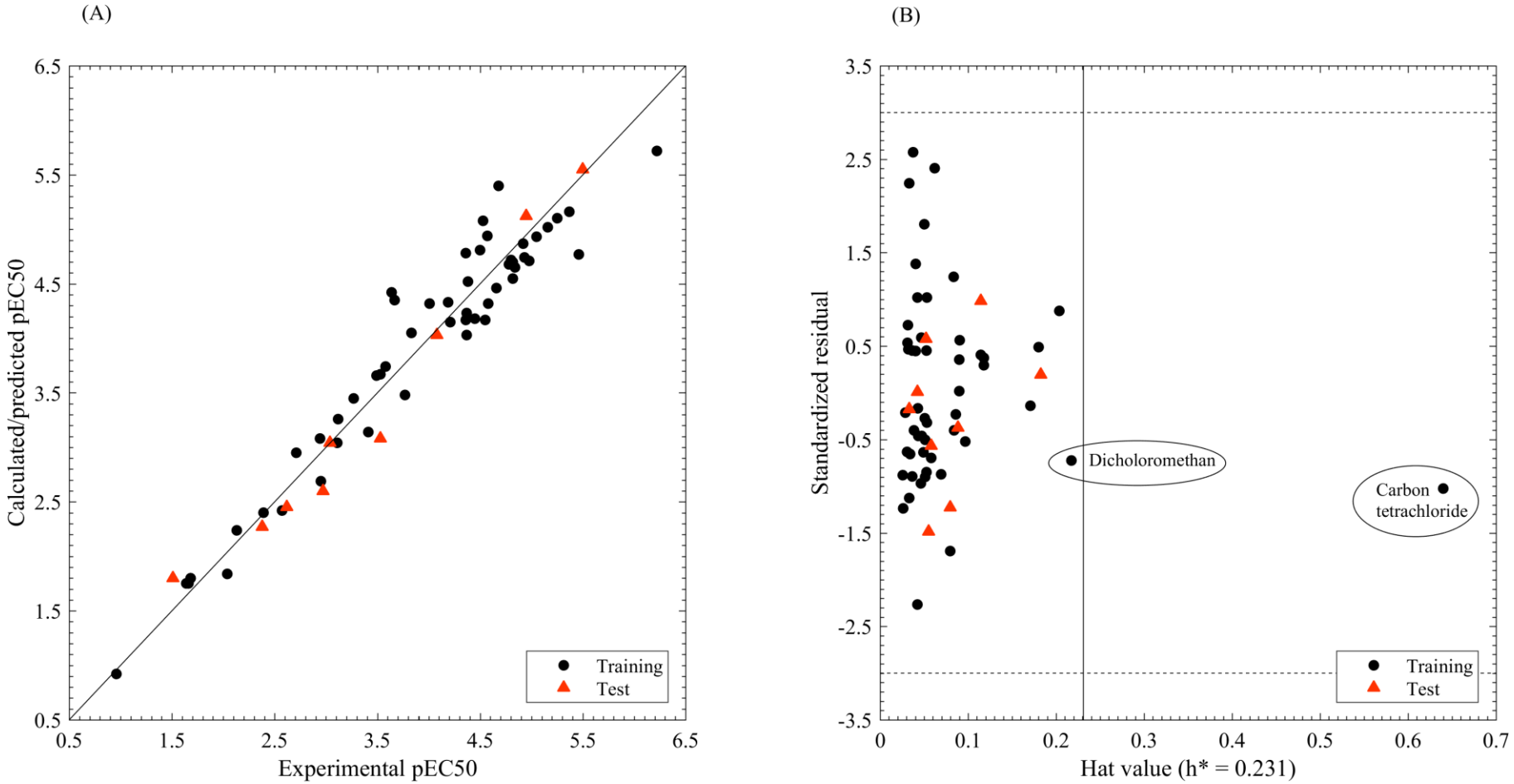

Figure 4.6. Predicted vs. experimental pEC_{50} (A), applicability domain (B) of the LM1.

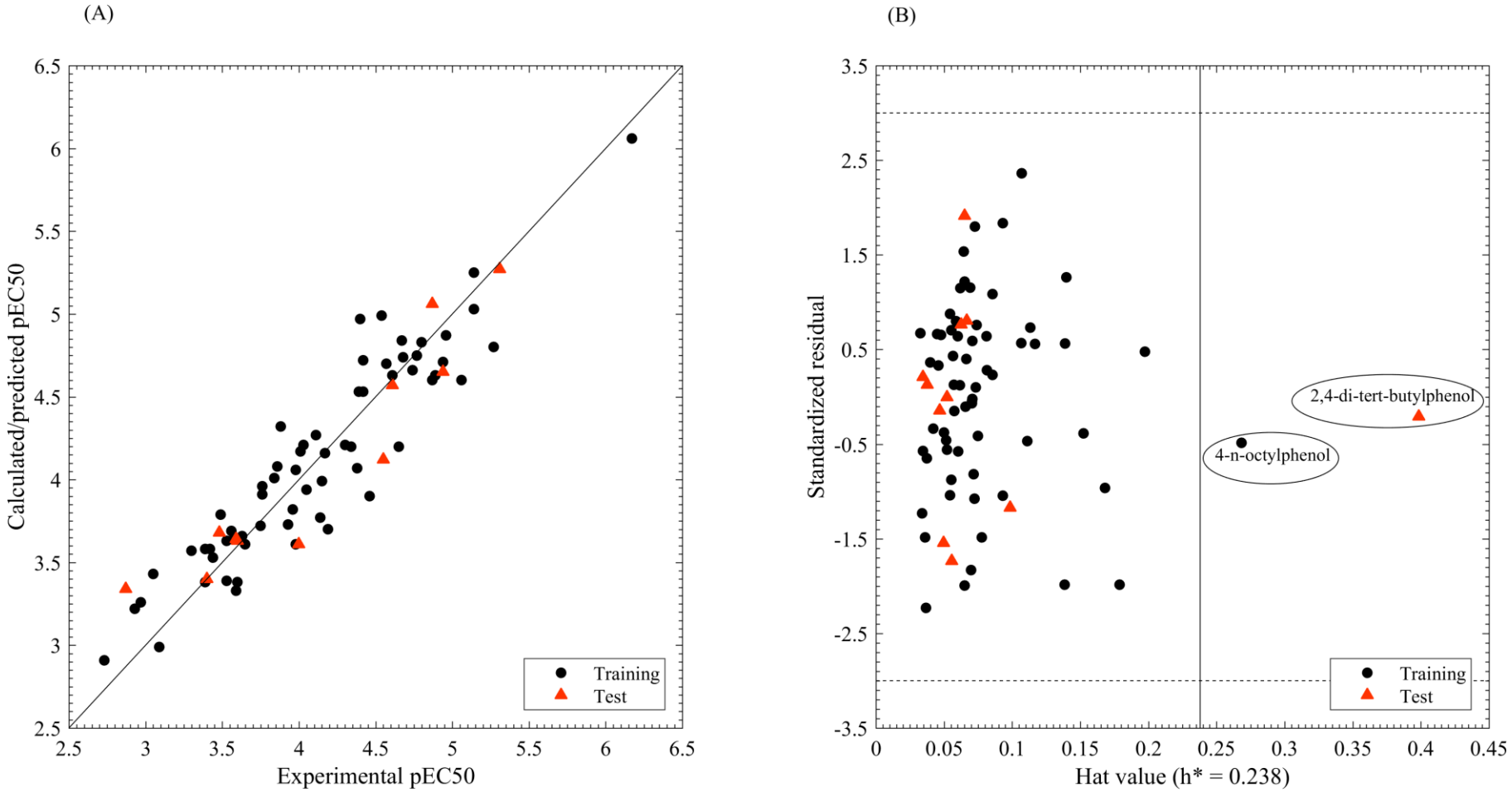

Figure 4.7. Predicted vs. experimental pEC_{50} (A), applicability domain (B) of the LM2.

Table 4.8. Linear and non-linear algae models from different studies.

Model	Species	Chemical class	Duration	Number of compounds	Number of descriptors	R^2	R^2_{TEST}	CCC_{TEST}	Reference
Linear models									
1	*P. subcapitata*	Substituted anilines and phenols	72-h	58	1	0.600	–	–	Aruoja et al., 2011
2	*P. subcapitata*	Benzo-(triazoles)	72-h	35	3	0.820	–	0.880	Gramatica et al., 2012
3	*P. subcapitata*	Propargylic alcohols	48-h	15	1	0.760	–	–	Chen et al., 2012
4	*C. vulgaris*	Phenols	96-h	30	2	0.820	0.640	–	Ertürk and Saçan, 2013
5	*P. subcapitata*	Non-polar and polar narcotics	72-h	108	3	0.915	0.924	–	Aruoja et al., 2014
6	*P. subcapitata*	Organic chemicals	48-h	105	5	0.770	0.800	–	Pramanik and Roy, 2014
7	*P. subcapitata*	Pharmaceuticals	72-h	45	4	0.790	–	0.850	Sangion and Gramatica, 2016
8	*C. vulgaris*	Phenols	96-h	46	3	0.860	0.940	0.970	Tugcu et al., 2017
LM1	Multispecies	Non-polar narcotics	72-h	61	3	0.931	0.967	0.981	**This study**
LM2	Multispecies	Polar narcotics	72-h	74	4	0.861	0.887	0.933	**This study**
GM	Multispecies	Diverse	72-h	455	8	0.661	0.718	0.837	**This study**
CPANN models									
9	*Duneliella tertiolecta*	Phenols	48-h	30	4	0.920	–	–	Ertürk et al., 2012
CPANN	Multispecies	Diverse	72-h	455	8	0.900	0.650	–	**This study**

Because a large proportion of the chemicals were not assigned to a certain MoA, log K_{ow} alone did not sufficiently explain the toxicity for the entire dataset. No significant difference between endpoint calculation methods and high interspecies correlations was reported by Fu et al. (2015). Elsewhere, strong correlations between different endpoints (Lessigiarska et al., 2004) and algal species (Ertürk and Saçan, 2012), along with QSTR models constructed for *C. vulgaris* but externally validated with *P. subcapitata* toxicity data (Ertürk and Saçan, 2013) supported the idea that modeling mixed algal toxicity data would be logical. These factors enabled searching for a global model (GM) based on the entire data.

Attempting to find the best (robust, validated, predictive, and the simplest) GM, several MLR models with different number of variables and training/test set divisions were generated. Structure-based division yielded better results. The 8-descriptor global QSTR model for the prediction of 72-h algal toxicity was selected following the procedure outlined in Section 3.5 (Equation 4.4).

$$\begin{aligned} pEC_{50} = {} & 5.140\ (\pm 0.473) + 3.484\ (\pm 0.710)\ \text{SPAM} + 1.924\ (\pm 0.251)\ \text{Mor31p} + 0.237\ (\pm 0.059)\ \text{NdsCH} \\ & - 0.439\ (\pm 0.093)\ \text{CATS2D\_02\_AP} + 0.950\ (\pm 0.151)\ \text{B05[C-S]} + 0.150\ (\pm 0.015)\ \text{F03[C-N]} \\ & + 0.098\ (\pm 0.007)\ \text{MLOGP2} - 0.765\ (\pm 0.074)\ \text{Hardness} \end{aligned} \quad (4.4)$$

Chemicals, experimental and predicted pEC_{50} values, final training/test set status, as well as calculated model descriptor values are presented in Table B.1. The GM is robust, satisfactory, and valid with respect to all the up-to-date rigorous statistical parameters (Table 4.6). The reliability of the model was confirmed by significantly low coefficients of R^2_{Yscr} and Q^2_{Yscr} following the Y-scrambling response randomization test. Moreover, high and close values for Q^2_{F1} and Q^2_{F2} indicated that the training set was representative and that the test set selection was homogenous in the response distribution. The good agreement between predicted and experimental data is presented by the homogenous distribution around the optimal line (Figure 4.8).

Type and chemical meanings of MLOGP2, hardness, F03[C-N], Mor31p, B05[C-S], SPAM, CATS2D_02_AP, and NdsCH (ordered by decreasing relative importance in Equation 4.4, based on standardized coefficients) are provided in Table 4.7.

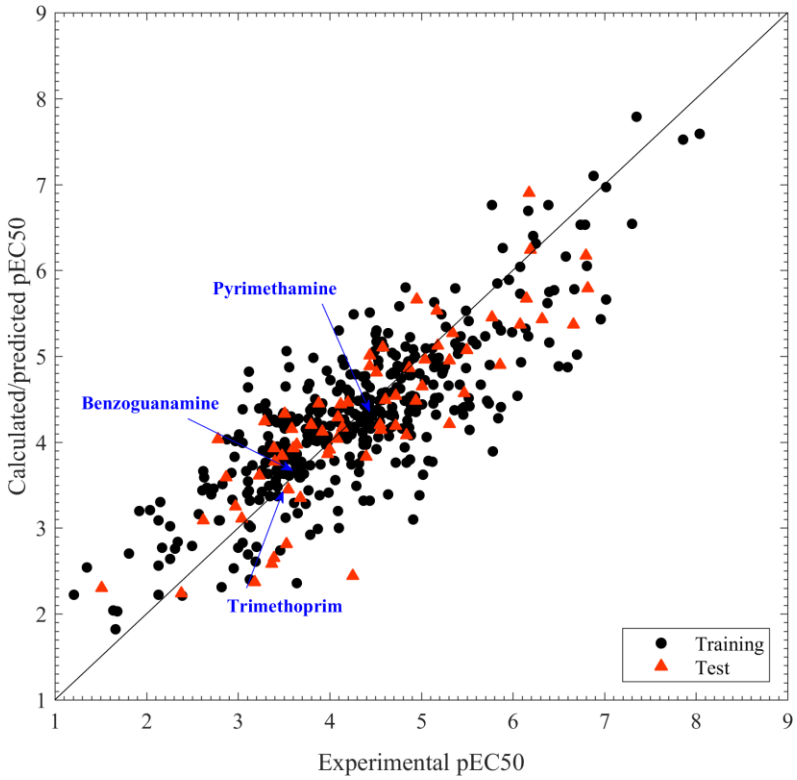

Figure 4.8. Predicted vs. experimental pEC_{50} of the GM.

All descriptors had positive contributions to toxicity, except for CATS2D_02_AP and hardness. The geometrical descriptor SPAM is calculated as the average value of conformational changes, indicating the flexibility of the molecule, and is used for describing long-chain molecules (Todeschini and Consonni, 2009). A positive SPAM coefficient indicates that the toxicity towards algae is directly proportional to the flexibility of the molecule. In a study by Pasha et al. (2007) SPAM appeared to be an important descriptor in a QSTR model developed for the toxicity of small organic molecules to *Tetrahymena pyriformis*. The descriptor Mor31p belongs to the 3D-MoRSE descriptors weighted by atomic polarizability. The information on atomic polarizability may be related to the alterations in the spatial arrangements of the substitution patterns within the electron diffraction properties. The electron distribution regarding the atomic polarizability of substituents might then be a factor influencing the toxic ability of the studied chemicals. Devinyak et al. (2014) reported a comprehensible way to interpret 3D-MoRSE descriptors in QSAR studies. The atom-type E-state indices descriptor NdsCH counts the number of unsaturated sp2 carbon atoms of the type =CH–. It describes a variety of functional groups with double-bonded carbon such as amides, aldehydes, carbamic and carboxylic acids, esters, ketones, and carbon-carbon double bonds. A typical characteristic of =CH– functionality is its tendency to give substitution or addition reactions because of the electrophilic nature of sp2

hybridized carbon. In the carbon-carbon double bond case, the p bond first acts as nucleophile, generating an electrophilic carbocation, which is then attacked by another nucleophile in additional reactions. The positive coefficient of NdsCH in Equation 4.4 shows that toxicity increases with the electrophilic character. The present results are in concordance with a study where NdsCH was reported to explain toxicity toward fathead minnow (Cassotti et al., 2015). The atom-pair descriptor CATS2D_02_AP is based on a 2D structure encoding topological information, where CATS refers to a chemically advanced template search and AP denotes a hydrogen bond acceptor, either positively charged or ionizable (Schneider et al., 1999). The negative contribution of this variable is in accordance with the above finding as the electronegativity of hydrogen bond acceptors is inversely proportional to electrophilicity. The descriptors B05[C-S] and F03[C-N] showed positive relationships to algal toxicity. A measure of hydrophobicity/lipophilicity, MLOGP2 is the squared Moriguchi K_{ow} (Moriguchi et al., 1992). Bearing a positive coefficient, MLOGP2 accounts for the increasing toxicity with increasing hydrophobicity, as expected, which was also reported to be an important descriptor in the fish toxicity modeling of pharmaceuticals (Tugcu et al., 2012). Hardness is a quantum chemical descriptor that provides information on the reactivity/stability of molecules. High values of hardness are related to the stability of a molecule (Todeschini and Consonni, 2009). The negative coefficient of hardness in Equation 4.4 is a clear expression of the inverse relationship between the stability and the reactivity of a molecule.

QSTR models are only functional in the domain in which they were trained and validated, regardless of the chemical diversity of the training data. Thus, reliable predictions are unlikely for substances outside the model domain. The AD of the GM was defined based on the leverage and the standardization approaches. The Williams plot (A) and a comparative analysis of the results based on both approaches (B) were presented in Figure 4.9, respectively. The structural outliers (enumerated compounds in (A)) that were not identified by the standardization approach are enclosed with an oval boundary in (B). Similarly, the standardization approach outliers (enumerated compounds excluding the ones enclosed with an oval boundary in (B)) that were not identified by the leverage approach are enclosed with a box boundary in (B).

Remarkably, the model covered all the chemicals in the response range. Similarly, no test set chemical was identified with hat and standardized descriptor values higher than the critical thresholds, suggesting that the toxicity predictions were reliably interpolated.

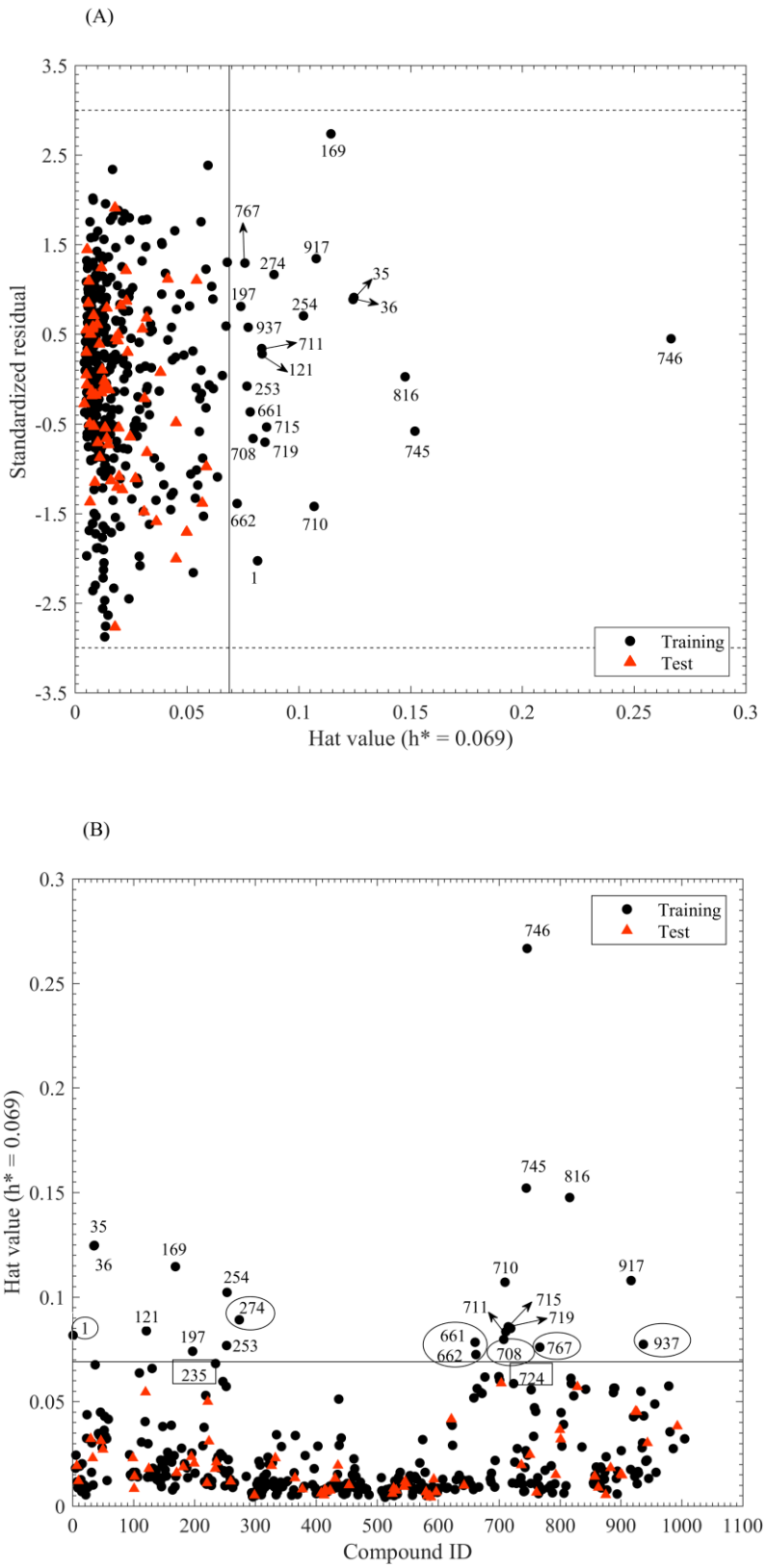

Figure 4.9. Applicability domain of the GM.

However, certain compounds of the training set were highlighted as outliers (numbering is the same as in Table B.1). The results showed that the same compounds generally appeared as outliers (35, 36, 121, 169, 197, 253, 254, 710, 711, 715, 719, 745, 746, 816, and 917) in both methods. Pyrimethamine (745), 2,4-diamino-6-phenyl-s- triazine (also known as benzoguanamine; 746), and trimethoprim (816) appeared as strong structural outliers, having high structural similarity and being the only compounds from the phenyldiamino diazine class out of 455 chemicals. Even so, as highlighted in Figure 4.8, these are "good leverage" compounds, lying perfectly along the regression line. In addition, 1,5-cyclooctadiene (35), 3a,4,7,7a-tetrahydro-1H-indene (36) (both belonging to the cyclo diene class), and sorbic acid (169) seemed structural outliers, all having the maximum value for NdsCH. Other high-leverage chemicals are as follows: terbumeton (708), 2-methylthio-4-tert-butylamino-6-amino-s-triazine (710), simetryn (711), terbutryn (715), and 2-methylthio-4-tert-butylamino-6-cyclopropylamino-s-triazine (719). These compounds belong to the alkylamino triazine chemical class and all had very high or maximum F03[C-N] values. 2-Tridecanone (121), 2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine) (197), and carbamazepine (917) have the highest Mor31p values; 1-decanethiol (254) and tetrabromobisphenol A (767) have the greatest MLOGP2, and cyclohexane (1) has the maximum hardness values. However, these high-leverage chemicals have no common structural features.

The model was applied to an external set of 320 industrial chemicals and pharmaceuticals with no experimental algal toxicity data (Table B.2). The external set covered a great majority of the chemicals addressed in "The List of Chemicals with no Ecotoxicological Data" (SU0303, 2015) announced by TÜBİTAK. Approximately one-fifth of the chemicals in the external set (61 of 320) are HPV chemicals, further implying the need for predicted toxicity values (Table B.2). The model provided high prediction capability and broad structural coverage (82%; only 57 of 320 external set chemicals appeared outside the AD). The prediction domain of the GM ranged from 1.82 to 7.79. It is worth noting that a few chemicals, such as itraconazole, a triazole antifungal, and azocyclotin, a pesticide with a triazole moiety appeared as both response and structural outliers (Figure 4.10). Other chemicals falling outside the structural AD are as follows: methotrexate (bears diazin), fusidic acid, prednisolone and methylprednisolone (bear polycycloalkanes), ethametsulfuron (bears triazin), bumetanide, cefuroxime, cefuroxime axetil, and novobiocin. Their predictions were extrapolated by the model.

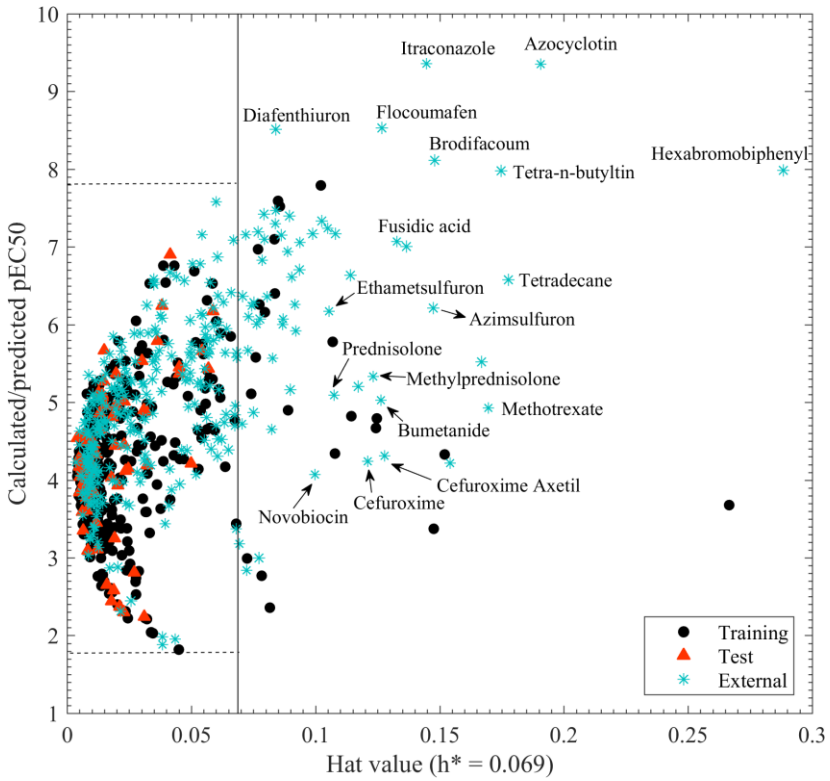

Figure 4.10. Prediction coverage of the GM.

Finally, the predicted pEC_{50} values of the external set chemicals were ranked in order to serve for screening and prioritization purposes in a scientific and regulatory frame. Accordingly, the most toxic 10 chemicals having the highest pEC_{50} values appeared as follows in the descending order: itraconazole, azocyclotin, flocoumafen, diafenthiuron, brodifacoum, hexabromobiphenyl, tetra-n-butyltin, triflupromazine, prometryn, and bupirimate. However, because some of them are outside the AD, the model is not adequate for these structures and possibly overestimated their toxicity. For a reliable screening, a new ranking within the AD led to the following priority list of the most toxic chemicals: triflupromazine, prometryn, bupirimate, carbosulfan, o,p'-DDT, ametryn, clethodim, pyroxsulam, dichlorodiphenyltrichloroethane (DDT), tetrasul.

A direct comparison of the global QSTR with recently published comparable studies (Basant et al., 2015; Singh et al., 2014; Villain et al., 2014) would not be appropriate due to different methodologies used. However, with an unambiguous algorithm and transparency, as well as reliability and validated statistical quality, the model provides practical advantages. Thus, it is potentially useful in regulatory decision-making and risk assessment.

4.2.2. Non-linear Model

Based on the assumption that MLR-based GM's descriptors encode significant information on toxicity, it was meaningful to make further efforts in order to understand how these descriptors would contribute to a non-linear model. Therefore, CPANN models were built using the same descriptors and training/test set division employed in MLR-based modeling for comparison purpose. The best model built with the network dimension of 21×21 and 700 epochs was selected following the procedure outlined in Section 3.5. A decent agreement between predicted and experimental data was observed by the homogenous distribution around the optimal line (Figure 4.11). CPANN model significantly outperformed the MLR model in the training set R^2, however, fell behind it in the test set (Table 4.6). The AD of the CPANN model was defined based on the EDs and displayed in Figure 4.12. Regarding the training set, only one chemical (cyclohexanone, ID 123) appeared outside the response domain. Compared to the MLR counterpart, CPANN predictions were closer to the experimental values for the training set. However, many test set chemicals were appeared outside the AD. These results might be due to the previously selected test set, implying that different training/test set divisions would have resulted in better statistical performance and less outliers. Nevertheless, considering that the main objective was to compare the two models, it can be concluded that satisfactory results were still obtained.

CPANN model was further challenged for the prediction of 72-h algal toxicity values of the external set. Predicted pEC_{50} values by the CPANN model for the dataset and external set are presented in Table B.1 and B.2, respectively. The model provided notable structural coverage (96%). The prediction range of the CPANN ranged from 1.82 to 7.38 and found comparable to the MLR counterpart. Only a few chemicals, such as azocyclotin, cefuroxime and cefuroxime axetil appeared as common outliers of both the MLR and CPANN models (Figure 4.13).

In order to compare, predicted pEC_{50} values by the MLR and CPANN-based models were mapped against each other revealing that both models were correlated in their predictive performance (r = 0.797 for the dataset, r = 0.713 for the external set). These results are presented in Figure 4.14.

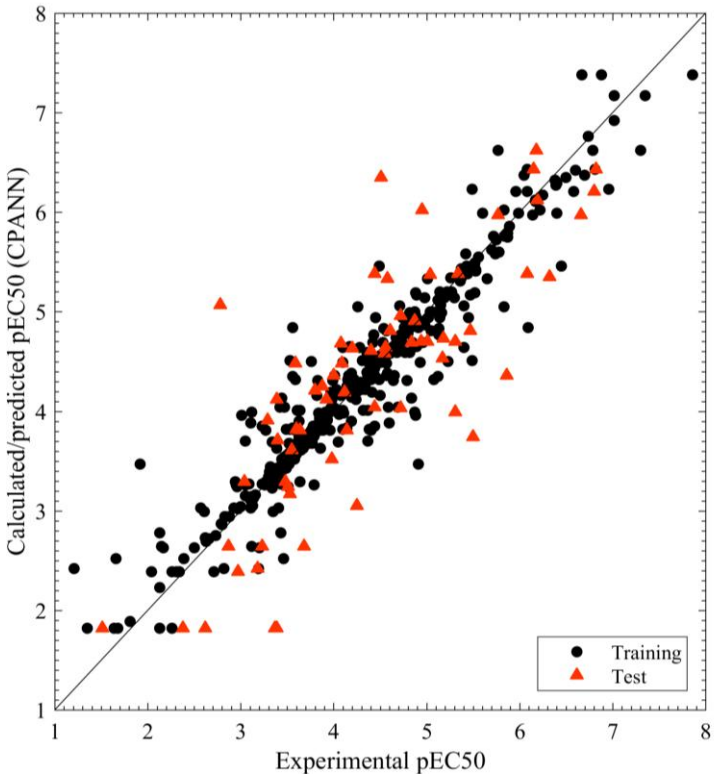

Figure 4.11. Predicted vs. experimental pEC_{50} of the CPANN model.

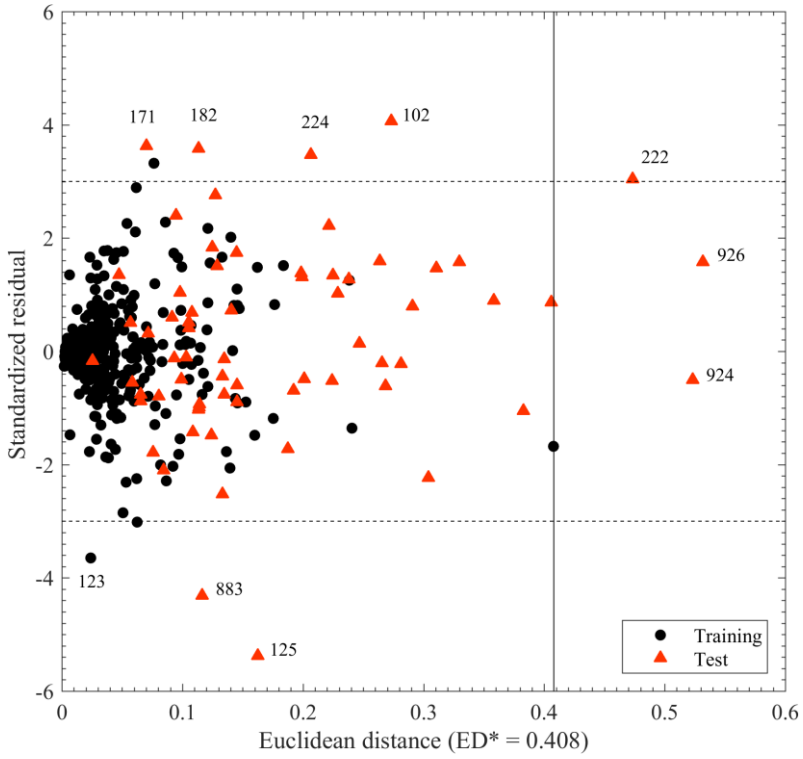

Figure 4.12. Applicability domain of the CPANN model.

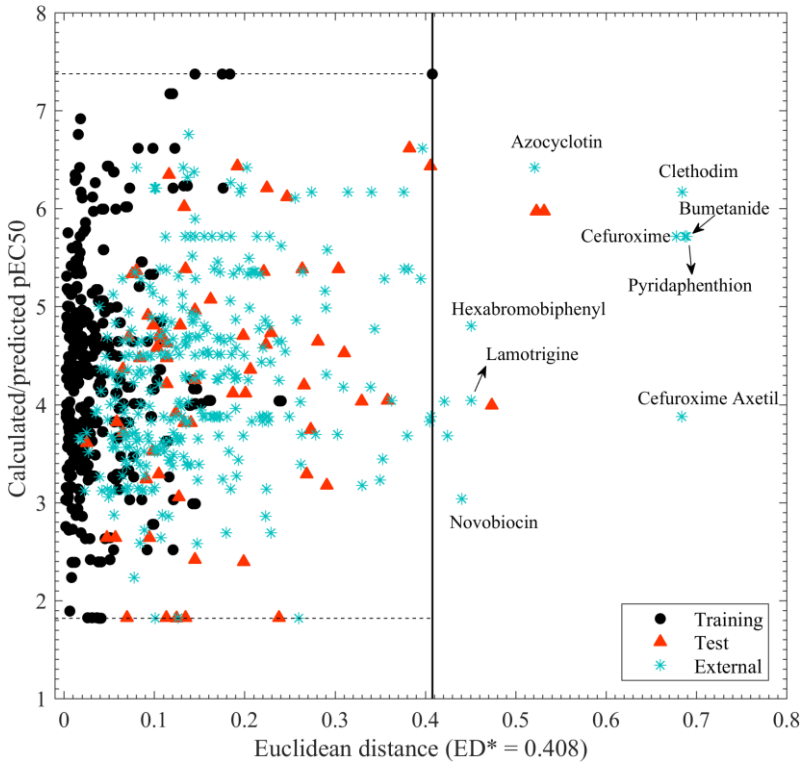

Figure 4.13. Prediction coverage of the CPANN model.

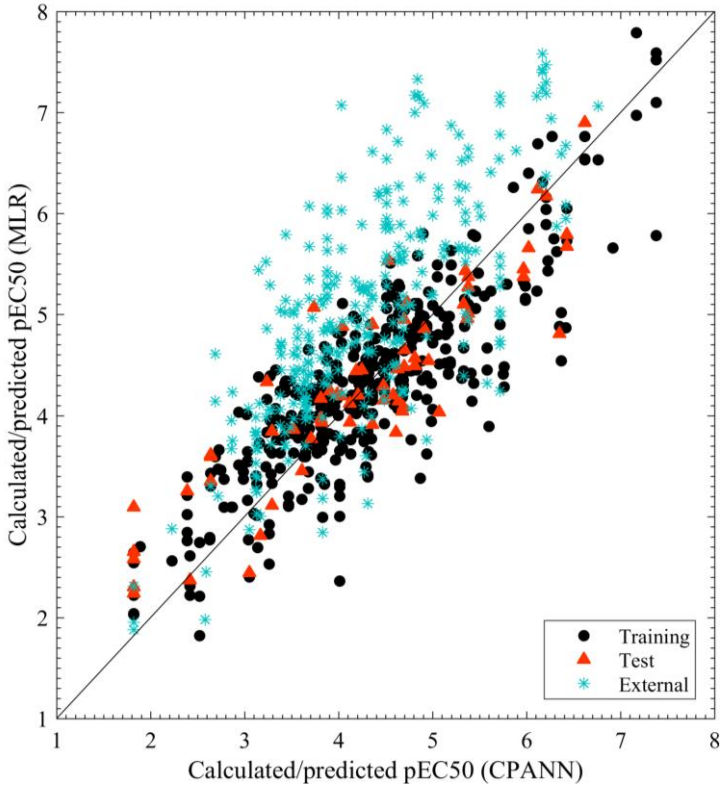

Figure 4.14. Predicted pEC_{50} by MLR vs. CPANN.

Finally, the influence of each descriptor on the modeled endpoint was analyzed by the graphical representation of the output layers (Figure 4.15). The visual analysis highlighted the good agreement between the weight maps of certain descriptors and toxicity. Among them SPAM, MLOGP2, and Mor31p displayed similar color distribution to that of pEC_{50}, revealing the positive contributions of these descriptors in encoding the toxicity. Likewise, having positive coefficients in the MLR model (Equation 4.4), these descriptors positively contribute to the modeled endpoint. On the other hand, the weights for CATS2D_02_AP and to some extent NdsCH showed dissimilar color distribution compared to the pEC_{50}. This result for CATS2D_02_AP was supported by the negative contribution of the same descriptor in the MLR-based model. Based on the slight red and strong blue tones in the output layers for B05[C-S] and F03[C-N], it can be concluded that these descriptors did not have significant influence on the pEC_{50}. It is important to note that the values of descriptors in the CPANN model were normalized between 0 (minimum value) and 1 (maximum value). The color distribution from red to blue for the output layer of pEC_{50} represents the corresponding maximum and minimum toxicity predicted by the CPANN model, respectively.

4.3. QSTR Models for RTL-W1 Cytotoxicity Prediction

The pEC_{50} values obtained from the AB and the CFDA-AM assays (Table 4.9) were used to model the RTL-W1 cytotoxicity. A high correlation (r = 0.986) was observed between the two cytotoxicity endpoints, $pEC_{50,\ CFDA\text{-}AM}$ and $pEC_{50,\ AB}$ (Figure 4.16). Similar results were previously reported in the review of Schirmer (2006). Therefore, it was anticipated that the same structural features (common descriptors) would encode both cytotoxicity endpoints.

In an attempt to find robust and validated models, several MLR models with different training/test sets combinations were generated for the endpoints obtained from the AB and the CFDA-AM assays. Identical descriptors (the number of aliphatic carboxylic acids (nRCOOH) and the highest occupied molecular orbital energy (E_{HOMO})) appeared in the models generated both for the $pEC_{50,\ AB}$ and $pEC_{50,\ CFDA\text{-}AM}$ endpoints. Chemicals, experimental and predicted $pEC_{50,\ AB}$ and $pEC_{50,\ CFDA\text{-}AM}$ values, final training/test set status, as well as calculated descriptor values are presented in Table 4.9. Both MLR models yielded nearly equivalent fit and prediction performances as well as their ADs were defined in the same way. Therefore, the detailed results of only one of the two models (AB-based model) were presented to avoid redundancy. Likewise, the $pEC_{50,\ AB}$ was referred to as the pEC_{50} for simplicity.

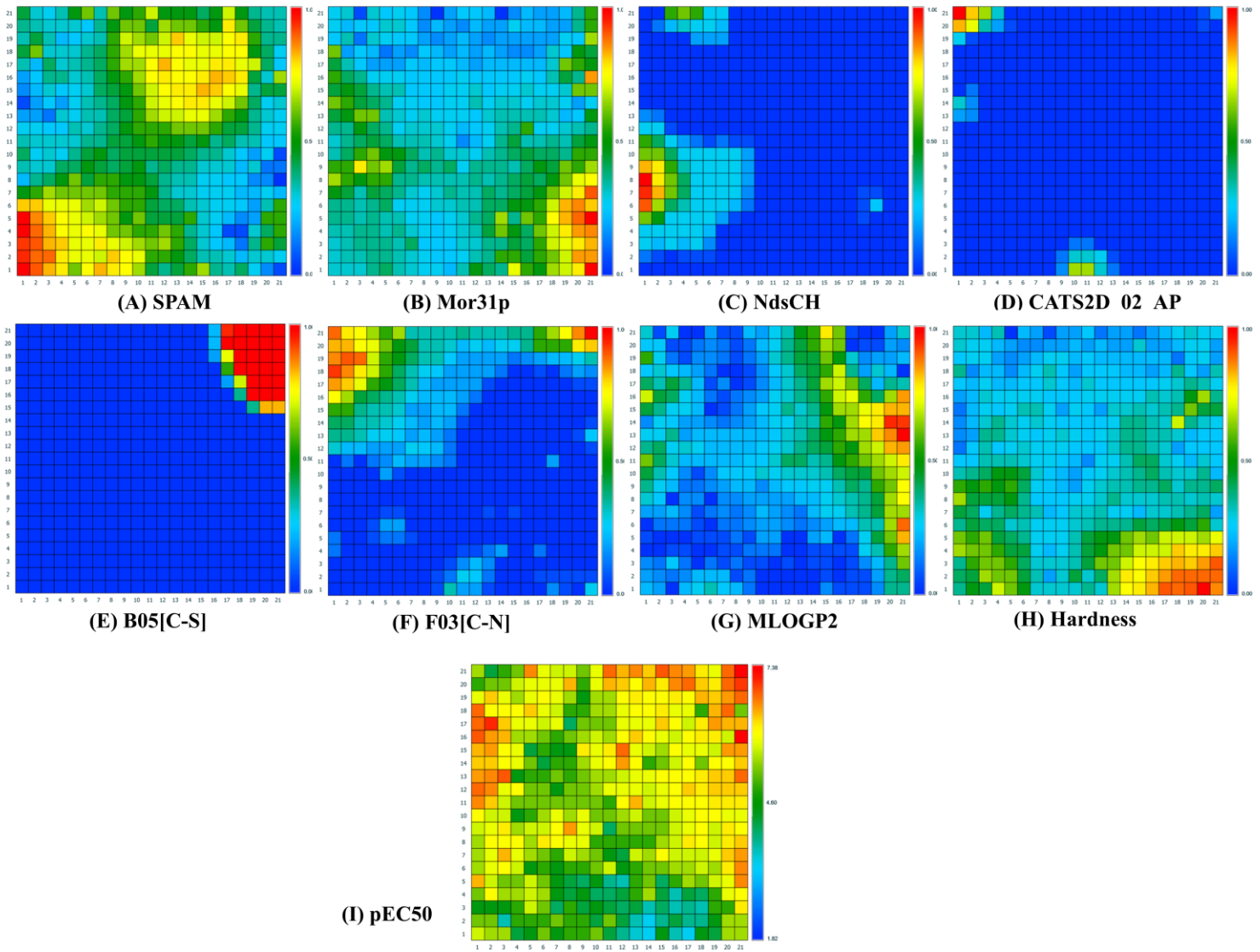

Figure 4.15. Output layers of the CPANN model.

Table 4.9. RTL-W1 dataset chemicals, experimental/predicted pEC_{50}, model descriptors.

CAS	Name	Exp $pEC_{50, AB}$	Pred $pEC_{50, AB}$	Exp $pEC_{50, CFDA-AM}$	Pred $pEC_{50, CFDA-AM}$	nRCOOH	E_{HOMO}
4474-24-2	Acid blue 80	-0.84	-0.75	-1.03	-0.81	0	-7.970
41859-67-0	Bezafibrate	-2.44	-2.50	-2.42	-2.48	1	-9.410
13171-00-1	Celestolide	-0.89	-1.25	-0.83	-1.29	0	-9.060
637-07-0	Clofibrate*	-1.34	-1.20	-1.39	-1.25	0	-8.970
15307-86-5	Diclofenac	-2.41	-2.08	-2.43	-2.07	1	-8.470
49562-28-9	Fenofibrate*	-1.45	-1.39	-1.32	-1.42	0	-9.380
54910-89-3	Fluoxetine*	-1.00	-1.33	-0.85	-1.37	0	-9.250
54739-18-3	Fluvoxamine	-1.66	-1.45	-1.70	-1.48	0	-9.510
1222-05-5	Galaxolide	-1.41	-1.14	-1.26	-1.18	0	-8.820
25812-30-0	Gemfibrozil*	-2.01	-2.11	-1.96	-2.10	1	-8.540
15687-27-1	Ibuprofen*	-2.41	-2.56	-2.32	-2.53	1	-9.540
22071-15-4	Ketoprofen	-2.30	-2.67	-2.20	-2.64	1	-9.790
81-14-1	Musk ketone	-2.42	-1.99	-2.32	-2.00	0	-10.700
81-15-2	Musk xylene	-2.14	-2.18	-2.32	-2.18	0	-11.130
22204-53-1	Naproxen	-2.34	-2.24	-2.36	-2.22	1	-8.830
61869-08-7/110429-35-1	Paroxetine	-0.92	-0.98	-1.04	-1.03	0	-8.470
87-86-5/131-52-2	Pentachlorophenol	-1.28	-1.44	-1.34	-1.48	0	-9.500
1506-02-1/21145-77-7	Tonalide	-0.86	-1.25	-0.89	-1.29	0	-9.060

*Test set compound.

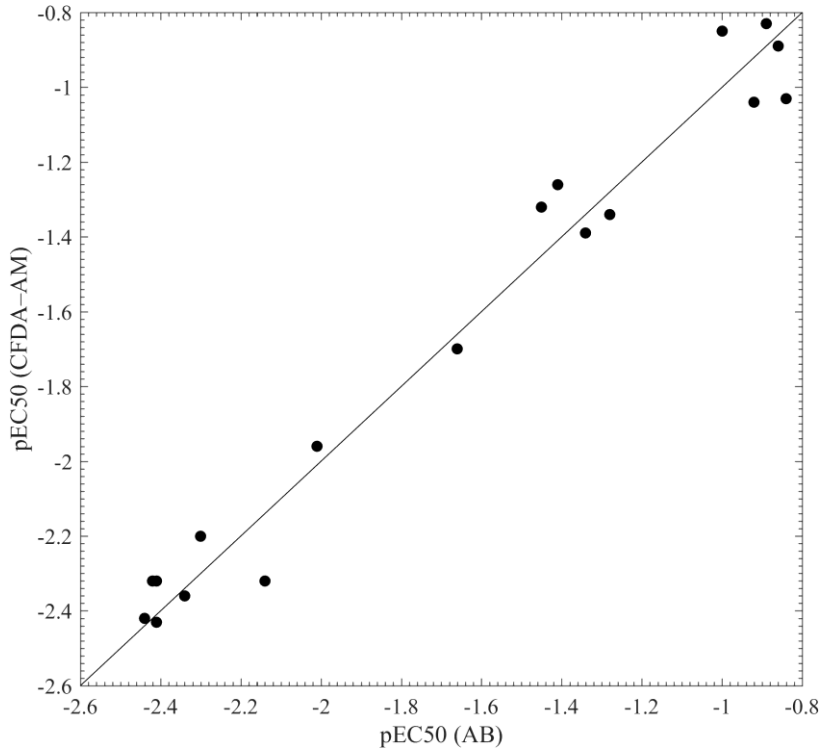

Figure 4.16. Experimental $pEC_{50,\ CFDA\text{-}AM}$ vs. $pEC_{50,\ AB}$.

The CFDA-AM model equation and related statistics are available as SM online at the article's web page https://onlinelibrary.wiley.com/doi/abs/10.1002/etc.3663 (Önlü and Saçan, 2017c).

AB-based linear QSTR model for the prediction of RTL-W1 cell line cytotoxicity is reported as Equation 4.5 together with the standard errors on the regression coefficients:

$$pEC_{50} = 2.859\ (\pm 0.923) - 1.098\ (\pm 0.180)\ nRCOOH + 0.453\ (\pm 0.098)\ E_{HOMO} \quad (4.5)$$

The two-descriptor model performance was found to be robust, validated and predictive, satisfying all the strict statistical acceptance criteria (Table 4.10). The fit of the model was verified by high R^2 and R^2_{adj} as 0.839 and 0.807, respectively. The internal validation parameter Q^2_{LOO} of 0.728 is comparable to R^2, implying the stability of the model. The reliability of the model was judged by Y-scrambling response randomization test (2000 scrambling iterations). Low coefficients of R^2_{Yscr} and Q^2_{Yscr} of 0.167 and −0.436, respectively, revealed that the model was not obtained by chance correlation. Furthermore, high and close values for Q^2_{F1} = 0.872 and Q^2_{F2} = 0.871 indicated that the test set selection was homogenous in terms of response distribution.

Table 4.10. Summary of RTL-W1 model, statistical parameters, literature thresholds.

Model[a]				**Fit and internal validation parameters**[b, c]						
n_{TR}/n_{TEST}	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}	R^2_{Yscr}	Q^2_{Yscr}
13/5	0.839[d]	0.807	0.261	0.913	26.055	0.728[d]	0.340	0.855	0.167	−0.436
External validation parameters[b, c]										
R^2_{TEST}	$RMSE_{TEST}$	MAE_{TEST}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2_m	Δr^2_m	k	$(R^2 - R_0^2) / R^2$
0.903[d]	0.181	0.155	0.872[d]	0.871[d]	0.923[d]	0.939[d]	0.860[d]	0.003[d]	0.951[d]	0.002[d]

[a]AB: Alamar Blue model (Eq. 4.5); n_{TR}, number of training and n_{TEST}, number of test set compounds. [b]R^2, coefficient of determination; R^2_{adj}, adjusted R^2; $RMSE_{TR}$ and $RMSE_{CV}$, root mean squared error (*RMSE)* of training set and cross-validation *RMSE*; CCC_{TR} and CCC_{CV}, concordance correlation coefficient (*CCC*) of training set and cross-validation *CCC*; *F*, Fisher statistics; Q^2_{LOO}, leave-one-out cross-validation correlation coefficient; R^2_{Yscr} and Q^2_{Yscr}, new coefficients following Y-scrambling procedure; R^2_{TEST}, coefficient of determination of test set; $RMSE_{TEST}$, *RMSE* of test set; MAE_{TEST}: mean absolute error of test set; CCC_{TEST}, *CCC* of test set. [c]Literature thresholds and references are explained in Section 2.3.8. [d]Parameters passing the thresholds.

The homogenous distribution of the data around the optimal line reflected a good agreement between predicted and experimental data (Figure 4.17 (A)).

The model was built with the following descriptors: nRCOOH and E_{HOMO}. No significant correlation was found between the two descriptors (r = 0.140). nRCOOH is a simple molecular descriptor whose entry represents the number of aliphatic carboxylic acid functionality of the molecule of interest. It is a 1D descriptor with no conformational dependency, therefore, can simply be derived from the recognized substructures within the molecule (Todeschini and Consonni, 2009). The negative coefficient of nRCOOH in Equation 4.5 reveals that cytotoxicity decreases with the increasing number of aliphatic carboxylic acid functional groups in a molecule. This finding is consistent with the structural properties of carboxylic acids. The hydrophilic and polar nature, as well as the hydrogen bond forming ability (due to both carbonyl −C=O and hydroxyl −OH moieties) make them more water-soluble, enhancing the metabolism and ultimately favoring elimination. E_{HOMO} is the highest orbital energy level containing electrons in the molecule, which gives information on the reactivity/stability of specific regions of molecules. Molecules with a high E_{HOMO} value can donate their electrons more easily compared to molecules with a low E_{HOMO} value, hence, are more reactive. E_{HOMO} is also a measure of the nucleophilicity of a molecule (Todeschini and Consonni, 2009). The positive coefficient implies that E_{HOMO} and cytotoxicity are directly proportional. Thus, cytotoxicity increases with increasing nucleophilicity and reactivity of the chemical.

The AD of the model was evaluated based on the leverage approach (Figure 4.17 (B)). Remarkably, no response and structural outlier were identified.

The model was applied to an external set of 836 chemicals with no experimental RTL-W1 cytotoxicity data (Table D.1). The external set addressed a great majority of the chemicals in“The List of Chemicals with no Ecotoxicological Data” (SU0303, 2015) announced by TÜBİTAK. The model provided remarkable prediction capability and structural coverage (93%; only 56 of 836 external set chemicals were outside the AD). The prediction domain of the model ranged from –2.67 to –0.75 for pEC_{50}, 0 to 1 for nROOH, and –11.13 to –7.97 for E_{HOMO}, respectively.

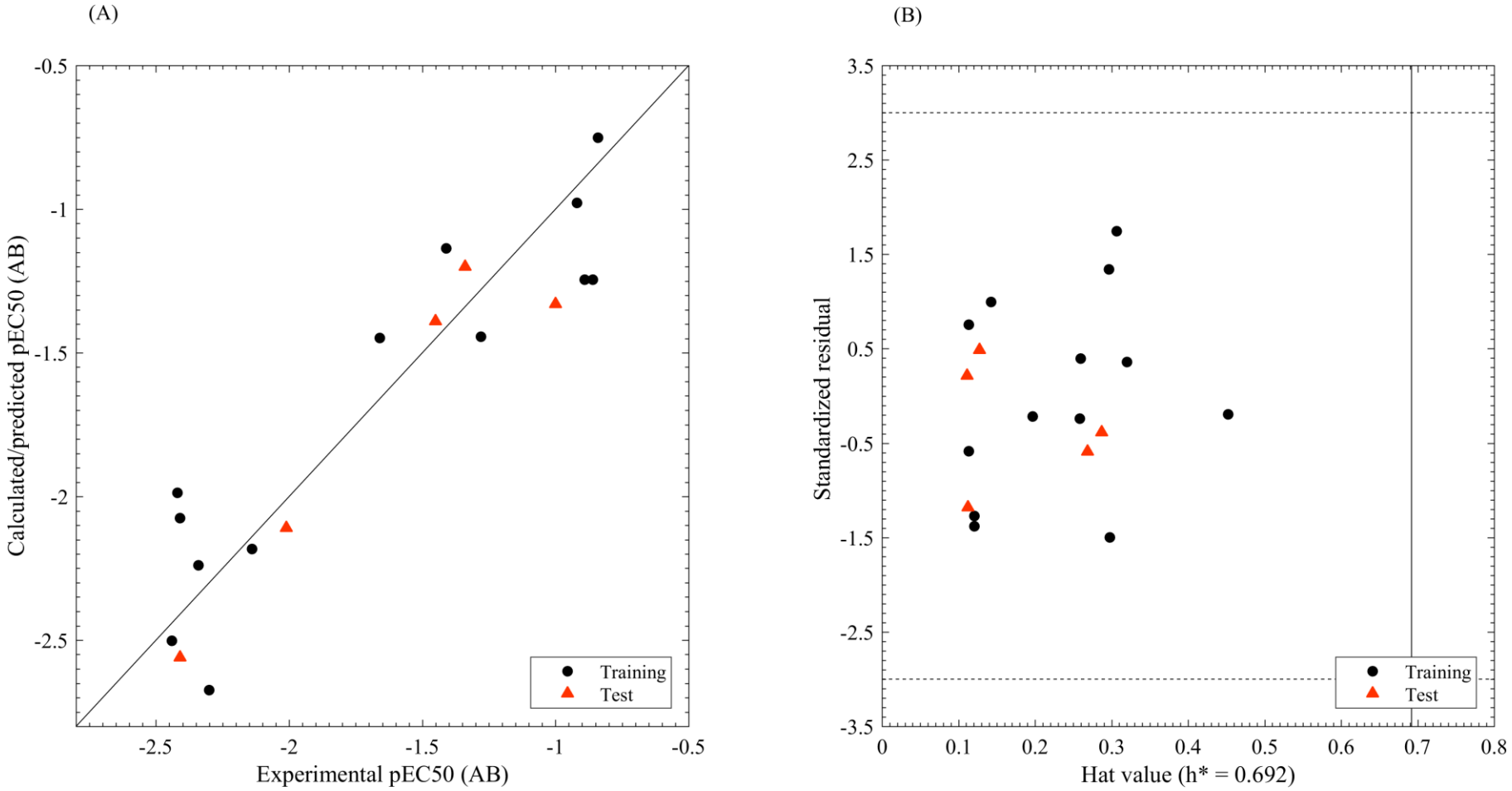

Figure 4.17. Predicted vs. experimental pEC_{50} (A), applicability domain (B) of RTL-W1 model.

It is important to note that oxalic acid, adipic acid, and perfluorooctanoic acid appeared as strong response and structural outliers (Figure 4.18). Oxalic acid and adipic acid are both dicarboxylic acids, i.e., with a nROOH = 2, which falls outside the descriptor domain of the model's training set. On the other hand, perfluorooctanoic acid is a bulky molecule with a lower E_{HOMO} value than the model's structural domain. Other significant outliers are as follows: methotrexate (bears diazin), folic acid (dicaarboxylic acid), acrylic acid (unsaturated carboxylic acid), perfluorooctane sulfonic acid, and 3,3'-thiodipropionic acid (dicarboxylic acid). Their predictions were extrapolated by the model.

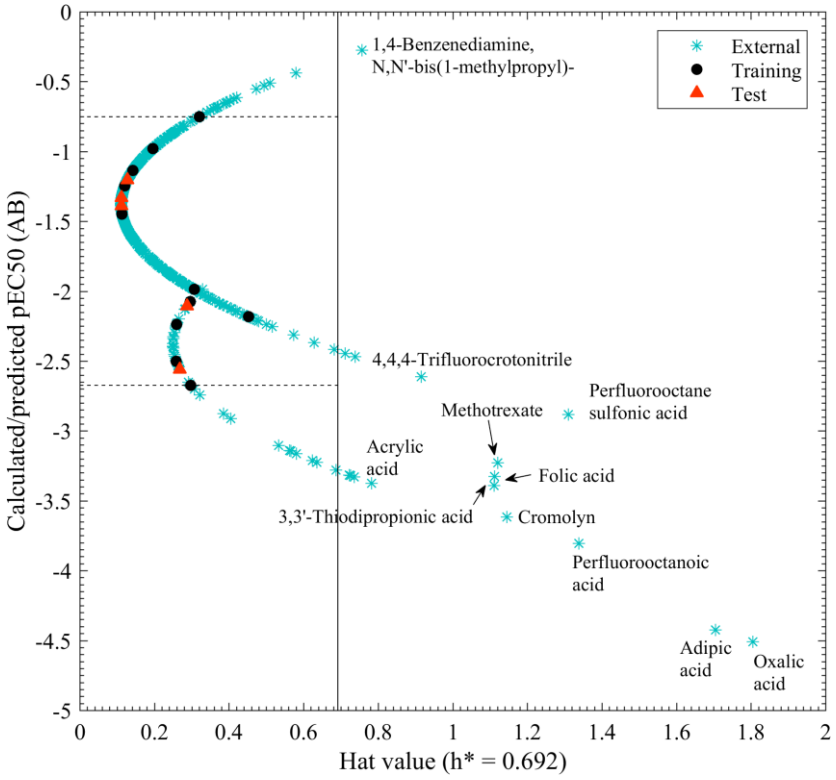

Figure 4.18. Prediction coverage of the RTL-W1 model.

The predicted pEC_{50} values of the external set chemicals within the model AD were ranked in order to provide information for screening and prioritization purposes. Accordingly, the most cytotoxic 10 chemicals having the highest pEC_{50} values appeared as follows in the descending order: 1,8-naphthylenediamine, 2,4-diaminotoluene, 1-(n-phenylamino)-naphthalene, 4,4'-diaminodiphenyl ether, p-anisidine, 2,4,6-trimethylaniline, n-phenyl-2-naphthylamine, terazosin, 2,4-dimethylaniline, and diphenylamine. No common chemicals appeared in the priority lists for 72-h algal toxicity and RTL-W1 cytotoxicity.

Another interest was to search for a possible relationship between *in vivo* fish acute toxicity and *in vitro* cytotoxicity. Therefore, *in vivo* LC_{50} values (mainly 96-h) available in the literature were collected for the dataset chemicals (data available as SM online at the article's web page https://onlinelibrary.wiley.com/doi/abs/10.1002/etc.3663). Of the 18 chemicals in the dataset, only 13 chemicals were found with a reported experimental pLC_{50}. A good agreement between experimental *in vivo* fish toxicity (pLC_{50}) and *in vitro* cytotoxicity (pEC_{50}) for the dataset chemicals was observed and presented in Figure 4.19.

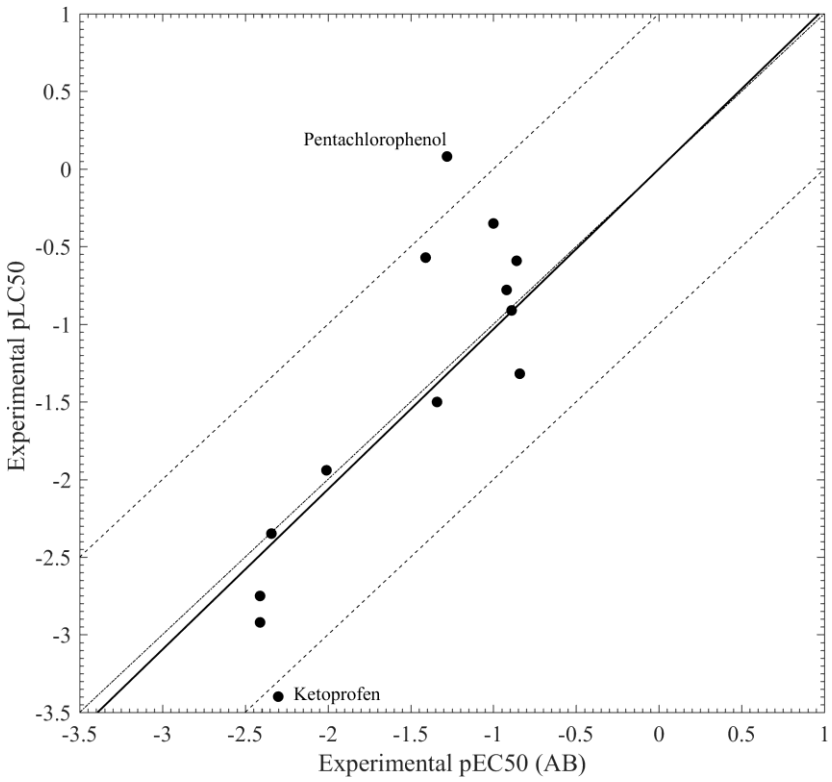

Figure 4.19. Experimental pLC_{50} vs. pEC_{50}.

In Figure 4.19, the solid line, dotted-dashed line, and dashed lines represent the fit line, the line of unity, and one order of magnitude deviation from the line of unity, respectively. The significant linear relationship (with a slope close to one) observed between *in vivo* and *in vitro* toxicity values is presented in Equation 4.6. However, pentachlorophenol and ketoprofen appeared as outliers. For pentachlorophenol, fish seemed to be about 20-fold more sensitive than fish liver cell line. Similarly, a good correlation was reported between fish gill cell line-based *in vitro* approach and acute fish toxicity data by Tanneberger et al. (2013). Although the pLC_{50} data is limited and contain structurally diverse

chemicals, it is worth noting that, Equation 4.6 can be used to predict the *in vivo* acute fish toxicity from experimental *in vitro* cytotoxicity which fall in the pLC_{50} range of .–3.40 to 0.08.

$$pLC_{50} = 1.032\ (\pm 0.106)\ pEC_{50} \quad (4.6)$$

$n = 13$, $R^2 = 0.887$, $R^2_{adj} = 0.878$, $F = 94.149$, $SE = 0.636$

Findings indicated that predicted RTL-W1 cytotoxicity data could be considered as an alternative to *in vivo* fish toxicity assays. Thus, it might be a useful method also for reducing the need for extensive *in vivo* testing. Finally, considering the need for fish acute toxicity data, pLC_{50} values for 836 external set chemicals were calculated from Equation 4.6, based on the predicted pEC_{50} values using Equation 4.5 (Table D.1). However, a cross-check of the calculated values against the available experimental values has not been carried out, which leaves an opportunity for future research.

4.4. QSTR and QTTR Models for *D. japonica* Toxicity Prediction

4.4.1. Hydrophobicity-Toxicity Relationship

Hydrophobicity is a universal physicochemical property elucidating cell membrane permeability, i.e., the initial step for eliciting toxic effects in aquatic organisms through narcosis. Based on the classical Hansch approach, first, a linear relationship between toxicity to *D. japonica* and hydrophobicity measured as log K_{ow} was investigated. Using experimental pLC_{50} and log K_{ow} for the 55 CEC (Table 4.11), a simple linear regression model for baseline toxicity to *D. japonica* was found (Equation 4.7):

$$pLC_{50} = -\ 3.052\ (\pm 0.147) + 0.395\ (\pm 0.046)\ \log K_{ow} \quad (4.7)$$

The standard errors of coefficients based on the OLS estimates are given in parenthesis. The moderate agreement observed between pLC_{50} and log K_{ow} (Table 4.12) suggests that even though it is an important parameter, hydrophobicity alone does not explain toxicity sufficiently. Besides, it is known that traditional Hansch approach applies to congeneric series of chemicals; yet, the dataset used in this study was structurally diverse. Moreover, from a MoA viewpoint, it is more likely that there are other mechanisms involved together with simple perturbation of membrane function.

Table 4.11. *D. japonica* dataset chemicals, experimental/predicted pLC_{50}, model descriptors.

ID	CAS	Name	Exp pLC_{50}[a]	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s	pEC_{50} (*D.magna*)[b]
1	104-40-5	4-nonylphenol*, §	-0.60	-0.63	-0.46	5.76	0.936	1.014	1	2.141	0.18
2	119-61-9	Benzophenone*, §	-1.44	-1.65	-2.01	3.18	1.125	1.000	0	-0.032	-1.62
3	131-56-6	2,4-dihydroxybenzophenone*	-1.12	-1.42		2.96**	1.093	0.995	2	0.985	
4	131-55-5	2,2',4,4'-tetrahydroxybenzophenone*	-1.42	-1.20		2.78**	1.083	0.990	4	1.584	
5	131-57-7	Oxybenzone*	-0.60	-1.45		3.79	1.067	0.984	0	1.354	
6	4065-45-6	Sulisobenzone*	-2.68	-2.60		0.37**	1.056	0.987	2	-1.537	
7	131-54-4	2,2'-dihydroxy-4,4'-dimethoxybenzophenone*	-1.71	-1.48		3.90**	1.054	0.976	0	1.393	
8	85-19-8	5-chloro-2-hydroxybenzophenone*	-0.84	-1.09		4.09**	1.170	1.006	0	1.435	
9	131-53-3	Dioxybenzone*	-1.26	-1.44		3.82**	1.070	0.982	0	1.427	
10	1641-17-4	Mexenone*	-0.79	-1.43		4.07**	0.974	0.983	1	0.979	
11	611-99-4	4,4'-dihydroxybenzophenone*	-1.76	-1.58		2.19**	1.083	0.999	4	-0.245	
12	117-99-7	2-hydroxybenzophenone†	-1.25	-1.35		3.52	1.109	0.996	0	1.693	
13	13020-57-0	3-hydroxybenzophenone*	-1.69	-1.71		2.67**	1.168	1.000	1	-0.503	
14	1137-42-4	4-hydroxybenzophenone†	-1.55	-1.51		3.07	1.101	0.999	2	-0.076	
15	1143-72-2	2,3,4-trihydroxybenzophenone*	-2.18	-1.25		2.91**	1.138	0.995	3	1.123	
16	99-76-3	Methyl 4-hydroxybenzoate*	-2.70	-2.92		1.96	0.653	0.953	0	-0.579	
17	120-47-8	Ethyl 4-hydroxybenzoate*	-2.27	-2.22		2.47	0.934	0.990	1	-1.379	
18	94-13-3	Propyl 4-hydroxybenzoate*	-1.83	-1.71		3.04	1.169	0.991	1	-0.856	
19	94-26-8	Butyl 4-hydroxybenzoate*	-1.60	-1.69		3.57	1.040	0.992	1	-0.772	
20	140-66-9	4-tert-octylphenol*, §	-0.67	-1.16	-0.31	5.28**	1.021	0.921	3	1.452	0.36
21	50-28-2	17β- estradiol*	-0.98	-1.16		4.01	0.862	0.996	4	0.944	
22	58-22-0	Testosterone†, ‡	-1.67	-1.40	-1.41	3.32	1.006	0.980	2	1.775	-0.92
23	57-63-6	17α-ethinylestradiol†, ‡	-0.76	-1.27	-1.72	3.67	0.868	0.996	4	0.824	-1.28
24	56-53-1	Diethylstilbestrol*, §	-0.42	-1.05	-1.14	5.07	0.917	0.944	6	0.090	-0.61
25	58-18-4	17α-methyltestosterone*	-1.62	-1.31		3.36	0.972	0.973	2	2.840	
26	13311-84-7	Flutamide*, ‡	-1.38	-1.00	-1.38	3.35	1.047	0.995	0	4.910	-0.88
27	80-05-7	Bisphenol A*, §	-1.56	-1.67	-1.90	3.32	0.926	0.966	4	-0.334	-1.49
28	882-09-7	Clofibric acid*, ‡	-2.79	-2.67	-2.80	2.57	0.604	0.965	1	-0.962	-2.53
29	134-62-3	N, N-diethyl-m-toluamide*, §	-2.88	-1.94	-2.85	2.18	1.264	0.940	0	1.558	-2.59
30	95-14-7	1H-benzotriazole*	-3.09	-3.07		1.44	0.000	1.047	0	0.212	
31	136-85-6	5-methyl-1H-benzotriazole*	-2.83	-3.04		1.71**	0.000	1.036	0	0.411	
32	103-90-2	Acetaminophen*, §	-3.39	-2.88	-2.48	0.46	0.783	1.001	0	-0.615	-2.16
33	50-78-2	Acetylsalicylic acid†, ‡	-2.73	-3.18	-2.93	1.19	0.535	0.979	0	-1.164	-2.69
34	15687-27-1	Ibuprofen†, §	-2.28	-1.72	-2.03	3.97	0.958	0.955	1	0.595	-1.64
35	61-68-7	Mefenamic acid*	-1.14	-0.82		5.12	1.360	0.977	2	-0.304	
36	22204-53-1	Naproxen*, §	-1.57	-1.71	-2.68	3.18	1.233	0.998	0	-1.127	-2.40

Table 4.11. Continued.

ID	CAS	Name	Exp pLC_{50}[a]	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s	pEC_{50} (*D.magna*)[b]
37	298-46-4	Carbamazepine*, §	-2.84	-2.23	-2.79	2.45	0.877	1.012	0	-1.221	-2.52
38	58-08-2	Caffeine*, ‡	-3.51	-3.70	-3.42	-0.07	0.004	1.021	0	-0.056	-3.25
39	57-62-5	Chlortetracycline*, §	-2.47	-2.47	-2.64	-0.62	0.815	0.973	5	1.483	-2.35
40	82419-36-1	Ofloxacin*, §	-2.73	-2.92	-2.44	-0.39	0.983	0.975	0	0.378	-2.12
41	723-46-6	Sulfamethoxazole*, §	-3.05	-2.75	-2.89	0.89	0.551	1.040	1	-1.277	-2.64
42	60-54-8	Tetracycline*, §	-3.23	-2.96	-3.32	-1.30	0.840	0.968	5	-0.502	-3.14
43	738-70-5	Trimethoprim†, ‡	-2.97	-3.27	-2.83	0.91	0.757	0.930	0	-0.730	-2.57
44	42200-33-9	Nadolol*, §	-3.30	-2.92	-2.88	0.81	0.673	0.923	2	1.292	-2.63
45	13523-86-9	Pindolol*	-2.43	-3.14		1.75	0.699	0.892	1	-0.033	
46	42399-41-7	Diltiazem*, §	-1.53	-2.04	-2.19	2.79**	1.051	0.943	0	1.226	-1.83
47	83881-51-0	Cetirizine*, §	-2.77	-2.69	-1.79	1.70	0.972	0.920	0	0.535	-1.36
48	58-73-1	Diphenhydramine*	-1.58	-1.91		3.27	1.007	0.947	0	1.362	
49	68-88-2	Hydroxyzine*	-1.43	-2.13		2.36**	0.988	0.962	0	1.071	
50	6138-79-0	Triprolidine*	-1.67	-1.36		4.89**	1.124	0.949	0	0.992	
51	79-06-1	Acrylamide†, §	-3.83	-4.53	-3.50	-0.67	0.000	0.946	0	-1.170	-3.35
52	2921-88-2	Chlorpyrifos*, §	-2.38	-1.83	-2.62	4.96	0.590	0.979	0	0.468	-2.32
53	62-73-7	Dichlorvos*	-1.08	-1.60		1.43	1.672	1.009	0	-0.950	
54	50-99-7	D-glucose*	-5.14	-5.23		-3.24	0.185	0.941	0	-2.315	
55	52645-53-1	Permethrin*	-2.81	-1.92		6.50	0.887	0.868	0	-0.404	

*****QSTR training set chemical. †QSTR test set chemical. §QTTR training set chemical. ‡QTTR test set chemical. ******Predicted log K_{ow} (EPI WSKOW v1.42). [a]Experimental pLC_{50} data of chemicals with the designated ID were taken from the following reference: 1 from Li, 2008; 2-19 from Li, 2012a; 20 from Li, 2012b; 21-31 from Li, 2013a; 32-50 from Li, 2013b; and 51-55 from Hagstrom et al., 2015. [b]pEC_{50} (*D. magna*) of chemicals with the designated ID were taken from the following reference: 1, 2, 20, 22, 26, 27, 29, 32, 34, 36-41, 44, 47, 51, 52 from Aalizadeh et al., 2017; 23, 24, 28, 33, 43, 46 from Sangion and Gramatica, 2016; and 42 from Li, 2013b.

Table 4.12. Summary of *D. japonica* models, statistical parameters, literature thresholds.

Fit and internal validation parameters[b, c]											
Model[a]	n_{TR}/n_{TEST}	**Descriptor(s)**	R^2	R^2_{adj}	$RMSE_{TR}$	CCC_{TR}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}	R^2_{Yscr}	Q^2_{Yscr}
Hydrophobicity	55/0	log K_{ow}	0.580	0.572	0.623	0.734	73.135	0.527	0.661	0.701	0.019	-0.057
QSTR	47/8	log K_{ow} GATS7p SpMaxA_G/D CATS2D_08_DL Mor31s	0.810[d]	0.787	0.419	0.895	34.964	0.731[d]	0.499	0.853	0.106	-0.187
QTTR	19/7	pEC_{50}	0.722[d]	0.706	0.520	0.839	44.121	0.666[d]	0.570	0.810	0.057	-0.189

External validation parameters[b, c]											
Model[a]	R^2_{TEST}	$RMSE_{TEST}$	MAE_{TEST}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{TEST}	r^2_m	Δr^2_m	k	$(R^2 - R_0^2) / R^2$
QSTR	0.891[d]	0.424	0.367	0.802[d]	0.799[d]	0.806[d]	0.919[d]	0.796[d]	0.099[d]	0.905[d]	0.030[d]
QTTR	0.839[d]	0.389	0.238	0.822[d]	0.822[d]	0.844[d]	0.895[d]	0.663[d]	0.167[d]	0.973[d]	0.015[d]

[a]Hydrophobicity model (Eq. 4.7); QSTR model (Eq. 4.8); QTTR model (Eq. 4.9); n_{TR}, number of training and n_{TEST}, number of test set compounds. [b]R^2, coefficient of determination; R^2_{adj}, adjusted R^2; $RMSE_{TR}$ and $RMSE_{CV}$, root mean squared error (*RMSE)* of training set and cross-validation *RMSE*; CCC_{TR} and CCC_{CV}, concordance correlation coefficient (*CCC*) of training set and cross-validation *CCC*; *F*, Fisher statistics; Q^2_{LOO}, leave-one-out cross-validation correlation coefficient; R^2_{Yscr} and Q^2_{Yscr}, new coefficients following Y-scrambling procedure; R^2_{TEST}, coefficient of determination of test set; $RMSE_{TEST}$, *RMSE* of test set; MAE_{TEST}: mean absolute error of test set; CCC_{TEST}, *CCC* of test set. [c]Literature thresholds and references are explained in Section 2.3.8. [d]Parameters passing the thresholds.

Therefore, it was worth investigating additional descriptors required to better relate toxicity with chemical structures. Experimental *D. japonica* pLC_{50} versus experimental/estimated log K_{ow} plot is available as SM online at the article's web page https://doi.org/10.1016/j.jhazmat.2018.02.046 (Önlü and Saçan, 2018).

4.4.2. QSTR for Predicting *D. japonica* Toxicity

Next, in an inquiry into finding a better relationship between chemical structures and pLC_{50}, a large variety of molecular descriptors in addition to log K_{ow} were utilized. Numerous log K_{ow}-based MLR models with different combination of variables using various training/test set divisions were developed. Overall, response-based division yielded better results. Based on the criteria outlined in Section 3.5, the best (robust, validated, predictive, and the simplest) model was eventually selected. 5-descriptor QSTR model for the prediction of *D. japonica* toxicity together with the standard errors of coefficients based on the OLS estimates is presented in Equation 4.8.

$$\begin{aligned} pLC_{50} = &- 10.415\ (\pm 1.861) + 0.279\ (\pm 0.040)\ \log K_{ow} + 1.132\ (\pm 0.219)\ \text{GATS7p} \\ &+ 6.604\ (\pm 1.865)\ \text{SpMaxA\_G/D} + 0.110\ (\pm 0.040)\ \text{CATS2D\_08\_DL} \\ &+ 0.147\ (\pm 0.055)\ \text{Mor31s} \end{aligned} \quad (4.8)$$

Chemicals, experimental and predicted pLC_{50} values, final training/test set status, as well as calculated model descriptor values are presented in Table 4.11. The results of the fit, internal, and external validation statistics are reported in Table 4.12. Of note, the addition of significant theoretical molecular descriptors brought additional structural information, leading to a better explanation of the relationship between chemical structure and toxicity. Thus, a remarkable improvement in the performance of the model was observed. The model meticulously meets the up-to-date statistical acceptance criteria (Table 4.12). The stability of the model was verified by Q^2_{LOO} comparable to R^2. Low values of R^2_{Yscr} and Q^2_{Yscr} confirmed the reliability of the model ensuring the absence of chance correlation. Likewise, close values of Q^2_{F1}, Q^2_{F2} and Q^2_{F3} revealed that the training set was representative and the test set selection was homogenous in the response distribution. The homogenous distribution around the optimal line reflects the good agreement between predicted and experimental toxicity (Figure 4.20 (A)).

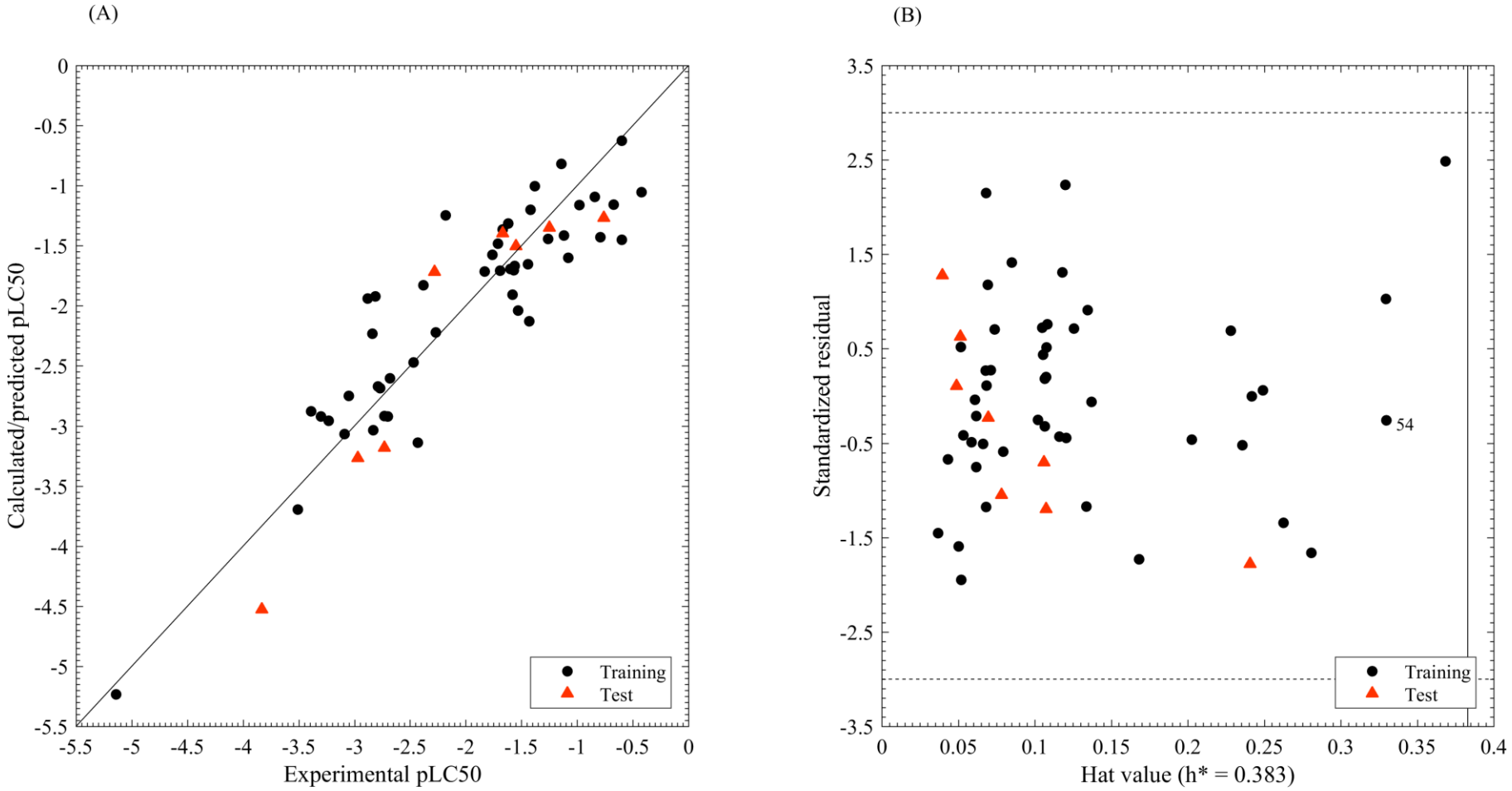

Figure 4.20. Predicted vs. experimental pLC_{50} (A), applicability domain (B) of QSTR model.

The model was built with the following descriptors ordered by decreasing relative importance in model equation based on the standardized coefficients: log K_{ow}, GATS7p, SpMaxA_G/D, Mor31s and CATS2D_08_DL. Bearing positive coefficients, all descriptors contribute to *D. japonica* toxicity positively. Log K_{ow} appeared as the most important parameter describing that the toxic action is due primarily to simple perturbation of membrane function. The second important descriptor, GATS7p, is a spatial 2D Geary autocorrelation descriptor of topological distance of lag 7 weighted by polarizability (Todeschini and Consonni, 2009). GATS7p encodes the atomic polarizabilities along the same topological distance. This suggests that the presence of chemicals with the aforementioned polarizability property might elicit an interaction resulting in toxicity. Similarly, GATS7p was reported as affecting the anti-malarial activity of synthetic prodiginines in another study (Masand et al., 2013). SpMaxA_G/D is a 3D matrix-based index for the folding degree of a molecule. It is the maximum eigenvalue of the geometric distance/topological distance quotient matrix (G/D) normalized by the number of atoms (Todeschini and Consonni, 2009). Its value approaches to 1 for linear molecules and decreases in response to the branching within the molecule, reflecting the changes in molecular size and shape, thus, flexibility. The positive coefficient supports that the higher SpMaxA_G/D is, the likelier the interaction leading to toxicity. Wang et al. (2015) reported that SpMaxA_G/D positively contributes to soil organic carbon normalized sorption coefficient. The descriptor Mor31s belongs to the signal-31 3D-MoRSE descriptors calculated upon the scattering parameter = 30 $Å^{-1}$ and weighted by the intrinsic state (I-state). The I-state of an atom can be interpreted as the possible partitioning of the effect of non-σ electrons throughout the σ bonds starting from the atom in question (Todeschini and Consonni, 2009). Hence, the less partitioning of the electron influence can be attributed to that the valance electrons are more prone to intermolecular interactions, which possibly result in toxicity. The last model parameter CATS2D_08_DL is a 2D structure-based atom-pair descriptor encoding topological information where CATS denotes a chemically advanced template search. DL represents a hydrogen bond donor-lipophilic pair of potential active centers (Schneider et al., 1999). The positive contribution of CATS2D_08_DL suggests that presence of a hydrogen bond donor and a lipophilic center at 8-bonds topological distance likely induces toxicity.

The AD of the QSTR model was evaluated based on the leverage and the standardization approaches (Figure 4.20 (B)). Remarkably, no response outlier was identified. This implies that the predicted toxicities of chemicals from the QSTR model are reliable. Likewise, no structural outlier was

found based on the leverage approach, whereas only one chemical; D-glucose (ID 54), was highlighted by the standardization approach.

4.4.3. QTTR for Predicting *D. japonica* Toxicity

Next, regarding the potential use of the existing experimental *D. magna* data for the estimation of *D. japonica* toxicity, the interspecies quantitative toxicity relationship was investigated. Earlier studies reported strong interspecies correlation between the experimental toxicity values for *D. japonica* and *D. magna* for limited data, suggesting that the studied chemicals might share similar toxicity mechanisms (Li, 2008; Li, 2012a; Li, 2012b; Li, 2013a; Li, 2013b; Liu et al., 2015). Of the 55 chemicals in the *D. japonica* dataset, only 26 chemicals with a reported experimental pEC_{50} data for *D. magna* were found (Table 4.11). A good agreement was found between the toxicity data of the two species (https://doi.org/10.1016/j.jhazmat.2018.02.046). Based on this, various QTTR models was developed using different response-based training/test set divisions. The best QTTR model for the prediction of *D. japonica* toxicity together with the standard errors of coefficients based on the OLS estimates is presented in Equation 4.9.

$$pLC_{50} = -0.620\ (\pm 0.277) + 0.860\ (\pm 0.130)\ pEC_{50} \qquad (4.9)$$

Likewise the QSTR model, the QTTR model also fulfilled the rigorous criteria regarding the fit, internal, and external validation statistics (Table 4.12). The good agreement between predicted and experimental toxicity was reflected by the homogenous distribution around the optimal line (Figure 4.21 (A)). Predicted pLC_{50} values and final training/test set status are presented in Table 4.11.

Regarding the AD of the QTTR model, noticeably, the standardization approach highlighted no outliers and only one chemical; 4-tert-octylphenol (ID 20), exhibited slightly higher leverage value than the critical one (Figure 4.21 (B)). Thus, it was identified as “good leverage” chemical reinforcing the model due to correct extrapolation, while extending the AD. Of note, 4-tert-octylphenol has the highest toxicity towards *D. magna*.

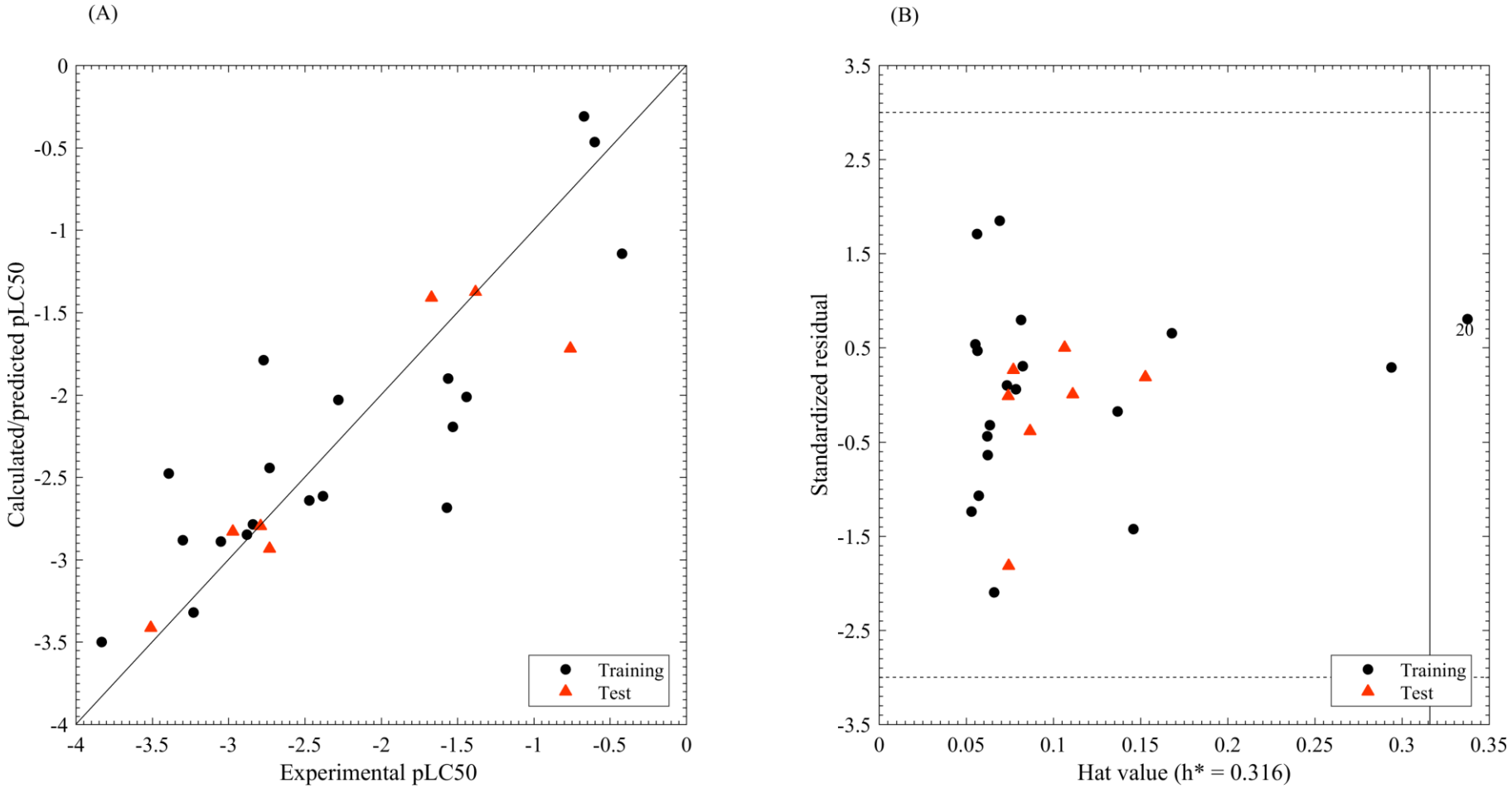

Figure 4.21. Predicted vs. experimental pLC_{50} (A), applicability domain (B) of QTTR model.

4.4.4. Application of QSTR and QTTR Models

The generated models were used to predict the toxicities of an external set of environmentally important chemicals with no experimental *D. japonica* toxicity data. The external set consisted of a wide variety of chemical classes, such as saturated and unsaturated hydrocarbons, aromatic alcohols, carbonyl compounds, and aniline derivatives, thus, was structurally heterogeneous. Predicted acute toxicity data for 792 industrial chemicals including many CEC and those in "The List of Chemicals with no Ecotoxicological Data" (SU0303, 2015) were reported for the first time (Table E.1). 317 of 792 chemicals (40%) were designated as HPV chemical emphasizing the importance of fulfilling the current data gap in regards to toxicity evaluation, screening and prioritization. With 677 chemicals in the AD, the QSTR model performed noticeable prediction coverage of 85% (Figure 4.22 (A)). However, 2,3-dichloro-1,4-naphthoquinone (ID 489) was distinctive as a strong structural outlier as well as its predicted pLC_{50} was outside the model prediction range. Regarding the application of the QTTR model, 266 of the 792 chemicals were found with a reported experimental pEC_{50} data for *D. magna* (Table E.1). Likewise, the QTTR model provided remarkable prediction coverage of 97%, having 259 chemicals within the AD (Figure 4.22 (B)). To interpret the reasons of the outliers, the ranges of response and descriptors of the QSTR and QTTR models were examined and presented in Table 4.13. For both models, the toxicity predictions of the chemicals outside the AD were due to extrapolation. However, it was not possible to identify shared structural characteristics for the outliers; therefore, no common structural features could be attributed. QSTR model descriptor ranges are as follows: log K_{ow}: −3.24 to 6.50, GATS7p: 0 to 1.672, SpMaxA_G/D: 0.868 to 1.047, Mor31s −2.315 to 4.910 and CATS2D_08_DL: 0 to 6, model pLC_{50} prediction range is −5.23 to −0.63. QTTR model descriptor range is as follows: pEC_{50}: −3.35 to 0.36, model pLC_{50} prediction range is −3.50 to −0.31.

In brief, both the QSTR and QTTR models are promising for being conspicuously simple and robust as well as complying with the stringent validation metrics and the OECD requirements (OECD, 2007). It is noteworthy to state that the models presented in this thesis allowed producing reliable information using the existing data, reducing the demand of *in vivo* and *in vitro* experiments on *D. japonica*, further contributing to the need for a more holistic approach to regulatory environmental safety assessment (Brown et al., 2016). Despite the heterogeneity in the *D. magna* toxicity data, the QTTR model performed satisfactorily, devoting to the protection of biodiversity.

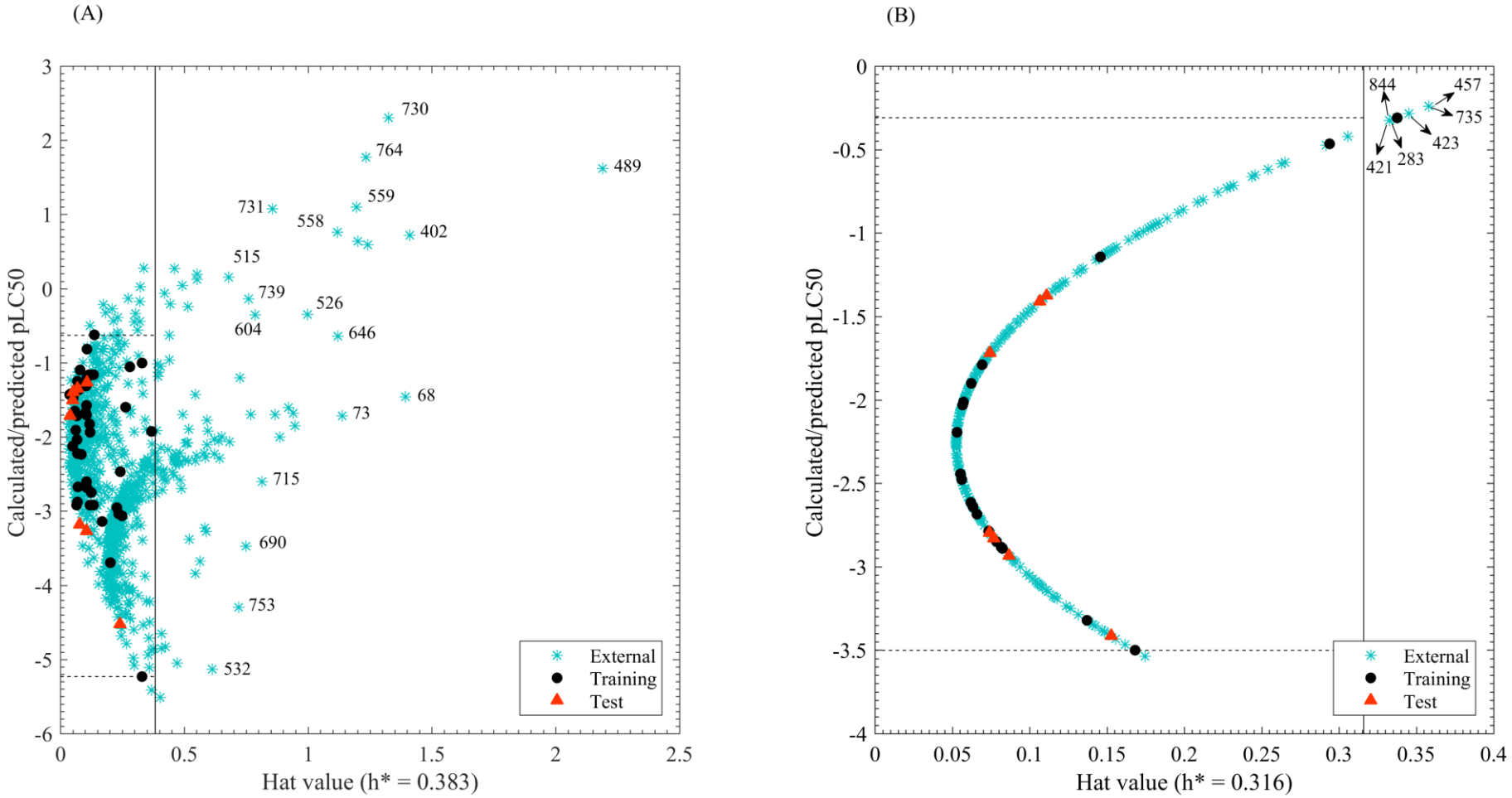

Figure 4.22. Prediction coverage of the QSTR (A) and QTTR (B) models.

Table 4.13. Reasons for extrapolated chemicals from QSTR and QTTR models.

QSTR Model			
ID	**Name**	**CAS**	**Reason**
68	Carbon tetrachloride	56-23-5	SpMaxA_G/D value higher than upper limit
73	Dibromochloromethane	124-48-1	SpMaxA_G/D value higher than upper limit
402	2,4,6-trichlorophenylhydrazine	5329-12-4	GATS7p value and predicted pLC_{50} value higher than upper limits
489	2,3-dichloro-1,4-naphthoquinone	117-80-6	GATS7p value and predicted pLC_{50} value higher than upper limits
515	6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	2439-01-2	GATS7p value and predicted pLC_{50} value higher than upper limits
526	Phenol,4,4',4"-ethylidynetris-	27955-94-8	CATS2D_08_DL value and predicted pLC_{50} value higher than upper limits
532	Methanol	67-56-1	Low predicted pLC_{50} value close to lower limit
558	3,4,5-trichloroguaiacol	57057-83-7	GATS7p value and predicted pLC_{50} value higher than upper limits
559	Tetrachloroguaiacol	2539-17-5	GATS7p value and predicted pLC_{50} value higher than upper limits
604	N-(3,5-dichlorophenyl)succinidide	24096-53-5	GATS7p value and predicted pLC_{50} value higher than upper limits
646	Acid blue 80	4474-24-2	CATS2D_08_DL value higher than upper limit
690	Ranitidine	66357-35-5	SpMaxA_G/D value lower than lower limit
715	Glibenclamide	10238-21-8	SpMaxA_G/D value lower than lower limit
730	Aldrin	309-00-2	GATS7p value and predicted pLC_{50} value higher than upper limits
731	Hexabromobiphenyl	36355-01-8	SpMaxA_G/D value and predicted pLC_{50} value higher than upper limits
739	Tetra-n-butyltin	1461-25-2	log K_{ow} value and predicted pLC_{50} value higher than upper limits
753	Folic acid	59-30-3	Mor31s value lower than lower limit
764	Endrin aldehyde	7421-93-4	Predicted pLC_{50} value higher than upper limit
QTTR Model			
ID	**Name**	**CAS**	**Reason**
283	4-n-octylphenol	1806-26-4	Predicted pLC_{50} and pEC_{50} values at higher limits
421	Atrazine	1912-24-9	High predicted pLC_{50} and pEC_{50} values close to higher limits
423	Terbumeton	33693-04-8	Predicted pLC_{50} and pEC_{50} values higher than upper limits
457	Oxyfluorfen	42874-03-3	Predicted pLC_{50} and pEC_{50} values higher than upper limits
735	Benzo[a]anthracene	56-55-3	Predicted pLC_{50} and pEC_{50} values higher than upper limits
844	Triallate	2303-17-5	pEC_{50} value higher than upper limit

Moreover, although the US EPA ICE application provides regression-based interspecies toxicity estimates for a few species using *Dugesia tigrina* as the surrogate (Dyer et al., 2006; US EPA, 2016), there is no such an option for *Dugesia-Daphnia* extrapolation to date. In addition to the rigorous OECD validation requirements, the QTTR model fulfills also the following guidance criteria for the inclusion of a regression model into the ICE application: relatively low MSE ($\leq$ 0.95), an adequate slope coefficient ($\geq$ 0.65), and a minimum sample size (Raimondo et al., 2015). Therefore, the QTTR model might contribute to this application as well.

QSTR-predicted pLC_{50} values of the external set chemicals within the model AD were ranked in an effort to generate information for screening and prioritization purposes. As a result, the most toxic 10 chemicals having the highest pLC_{50} values appeared as follows in the descending order: acid blue 80, N,N-dimethyldodecan-1-amine, undecane, terbutryn, perfluorooctanoic acid, diafenthiuron, atovaquone, methyl dodecanoate, 2,4-di-tert-butylphenol, and 4-n-octylphenol. No common chemicals appeared in the priority lists for 72-h algal toxicity and RTL-W1 cytotoxicity and *D. japonica* acute toxicity.

Finally, the same screening was applied separately on the "The List of Chemicals with no Ecotoxicological Data" to identify the common chemicals showing the highest and the lowest toxicity to algae, *D. japonica* and RTL-W1. Accordingly, the most and least toxic 10 chemicals were reported, respectively (Table 4.14). Carbosulfon, an insecticide, appeared among the most toxic chemicals to algae and RTL-W1, whereas dichlorodiphenyltrichloroethane (DDT), a formerly used pesticide, and N,N-dimethyldodecan-1-amine were among the most toxic chemicals to algae and *D. japonica*. Similarly, diafenthiuron was one of the most toxic chemicals to both *D. japonica* and RTL-W1. On the other hand, tetraethyl pyrophosphate, trichlorphon (chlorphos), hymexazol, and formothion were the least toxic to algae and *D. japonica*, while n-hexane and 2,2-dibromo-2-cyanoacetamide appeared as the least toxic to both algae and RTL-W1. Tetraethyl pyrophosphate, an organophosphate insecticide, appeared as the least toxic to all three species. Interestingly undecane was the most toxic to *D. japonica* but the least to RTL-W1.

Table 4.14. The most and the least toxic chemicals with no ecotoxicological data.

The most toxic 10 chemicals[a]					
Algae/Growth inhibition		***D. japonica*/Acute toxicity**		**RTL-W1/Cytotoxicity**	
Chemical[b]	**pEC_{50} (mol/L)**	**Chemical**	**pLC_{50} (mol/L)**	**Chemical**	**pLC_{50} (mol/L)**
Bupirimate	7.42	*N,N-dimethyldodecan-1-amine*	*5.35*	Methiocarb sulfoxide	5.14
Carbosulfan	*7.40*	***Undecane***	***5.33***	Fenamiphos sulfoxide	5.12
o,p'-DDT	7.33	*Diafenthiuron*	*5.26*	Ethiofencarb sulfoxide	5.10
Pyroxsulam	7.19	9,10-anthracenedione	5.06	*Carbosulfan*	*5.10*
Dichlorodiphenyltrichloroethane (DDT)	*7.17*	*Dichlorodiphenyltrichloroethane (DDT)*	*5.02*	Carbazole	5.08
Tetrasul	7.17	1,3-diphenylbenzene	5.00	*Diafenthiuron*	*5.08*
Aldrin	7.15	Tralkoxydim	4.99	Fenthion sulfoxide	5.06
1,2,4-trichloro-5-(3,4-dichlorophenyl)benzene	7.00	N,N-dimethyltetradecylamine N-oxide	4.95	Etofenprox	5.06
Metaflumizone	6.87	Fluorochloridone	4.89	2,2'-dichloro-4,4'-methylendianiline	5.04
N,N-dimethyldodecan-1-amine	*6.71*	Tau-fluvalinate	4.88	Benzo[a]anthracene	5.04
The least toxic 10 chemicals[a]					
Algae/Growth inhibition		***D. japonica*/Acute toxicity**		**RTL-W1/Cytotoxicity**	
Chemical	**pEC_{50} (mol/L)**	**Chemical**	**pLC_{50} (mol/L)**	**Chemical**	**pLC_{50} (mol/L)**
Metaldehyde	2.45	Folic acid	1.70	Mecoprop-P	3.49
Tetraethyl pyrophosphate	*2.88*	*Tetraethyl pyrophosphate*	*1.79*	4-chlorophenoxyacetic acid	3.60
n-hexane	*3.00*	Maleic hydrazide	1.79	HCH-delta	3.79
Trichlorphon (Chlorphos)	*3.40*	*Hymexazol*	*2.23*	2-naphthyloxyacetic acid	3.80
Hymexazol	*3.60*	Cis-1,2,3,6-Tetrahydrophthalimide	2.39	*2,2-dibromo-2-cyanoacetamide*	*3.80*
(4-chlorophenyl)urea	3.67	*Trichlorphon (Chlorphos)*	*2.53*	*Tetraethyl pyrophosphate*	*3.81*
2,2-dibromo-2-cyanoacetamide	3.67	Metalaxyl-M	2.53	HCH-alpha	3.83
3-hydroxycarbofuran	3.70	*Formothion*	*2.56*	*n-hexane*	*3.88*
Methiocarb sulfone	3.75	Aldicarb-sulfoxide	2.69	Tritosulfuron	3.91
Formothion	*3.76*	Aldicarb-sulfone	2.73	***Undecane***	***3.95***

[a]Based on the predicted values from Eq. 4.4 for algae, Eq. 4.5 for RTL-W1, and Eq. 4.8 for *D. japonica*. Toxicity values reported in the same unit for comparison. [b]Common chemicals appearing as the most toxic to two species are in italics. Common chemicals appearing as the most toxic to one species and the least toxic to another species are in bold and italics. Common chemicals appearing as the least toxic to all species are in italics and underlined.

5. CONCLUSIONS

Quantitative effect of different geometry optimization methods (semi-empirical, *ab-initio* and density functional theory) on quantum chemical and DRAGON-derived descriptors were systemically analyzed based on structurally diverse chemicals. A comparative reference summary for optimal method selection was presented. Using the knowledge gained from this analysis, binding affinity to human serum albumin was modeled as a case study, for the first time, by rationally selecting the geometry optimization method. Results provided evidence that the geometry optimization method significantly affects certain descriptors, thus, the statistical quality and performance of QSAR models built with descriptors sensitive to the geometry optimization method. An activity-independent rational approach towards selecting an optimal geometry optimization method for improved QSAR/QSTR modeling was proposed for the first time. The findings have contributed to the scientific field by providing an understanding of the impact of quantum chemical calculation methods on the estimation of molecular descriptors and model performance.

Available ecotoxicity data on representative aquatic species (algae, aquatic invertebrates, fish) for hundreds of environmentally significant chemicals were comprehensively analyzed. Data gaps were determined in a scientific and regulatory framework.

Validated QSTR models based on multiple linear regression (MLR) and counter propagation artificial neural network (CPANN) were developed for acute toxicity prediction towards mixed algae species (predominantly *Pseudokirchneriella subcapitata*) using the rational approach presented. Robust and predictive QSTR models were built over a large data of structurally diverse industrial chemicals. Results were found comparable to the existing studies in the field. The global QSTR model has contributed to the literature by providing wider applicability domain and notable prediction coverage.

The first report on the cytotoxicity prediction to rainbow trout liver cell line (RTL-W1) was proposed. A validated MLR-based QSTR model was developed based on the *in vitro* cytotoxicity data of structurally heterogenous pharmaceuticals and personal care products (PPCPs). High prediction ability and wide applicability domain were characterized for the model. The possible use of cytotoxicity values for the estimation of acute fish toxicity was evaluated. A significant linear

relationship was generated between experimental *in vitro* cytotoxicity and *in vivo* fish acute toxicity based on the available data. Results indicated that the integrative use of the cytotoxicity QSTR model with the linear *in vitro-in vivo* relationship is promising both for the prediction of acute toxicity of PPCPs on nontarget organisms as a first screening tool and reducing the need for extensive *in vivo* testing.

A validated linear QSTR model was presented to estimate *a priori* the acute toxicity of chemicals with emerging environmental concern towards *Dugesia japonica*, for the first time. Moreover, interspecies quantitative toxicity relationship between *Daphnia magna* and *D. japonica* was investigated and a validated QTTR model was reported. The QSTR and QTTR models enabled producing reliable information using existing data, while reducing the need for *in vivo* and *in vitro* experiments and devoting to the protection of biodiversity, further addressing a more holistic approach to environmental safety assessment. Both models were found promising for being noticeably simple and robust. The novel QTTR model might also contribute to the US EPA Interspecies Correlation Estimation web application.

Finally, the developed models were applied on hundreds of chemicals with no toxicity and cytotoxicity data. Consequently, predicted toxicity data to mixed algae species and *D. japonica*, as well as predicted cytotoxicity data to RTL-W1 were reported for a wide range of industrial chemicals including high production volume chemicals. Predicted toxicity and cytotoxicity data were reported for a great majority of the chemicals addressed in “The List of Chemicals with no Ecotoxicological Data” announced by the Scientific and Technological Research Council of Turkey (TÜBİTAK) for the first time. Based on the predicted toxicity and cytotoxicity values, a further screening allowed identifying the most and least toxic chemicals to each of the aquatic species studied in the present study.

In conclusion, validated alternative methods were presented for ecotoxicity assessment, screening, and prioritization of chemicals in a scientific and regulatory frame. The data gaps in these fields were filled substantially.

Regarding the future perpectives, the outcome of this study could lead up to the following topics:

Toxicity mechanism is a complex issue, therefore, non-linear modeling other than CPANN could be carried out for large toxicity datasets. Moreover, regarding the complexity of mechanistic interpretation of toxicological events of chemicals in living organisms, the integration of ligand-based (i.e., linear and non-linear QSTR) and structure-based (i.e., toxicophore modeling and molecular docking) approaches could be of utmost importance to assess the possible interactions between chemicals and target molecules. Risks resulting from the unintentional co-occurrence of chemicals in real environment, in other words, mixture toxicity could be a matter of concern. Consequently, mixture toxicity data are needed not only to understand the adverse effects of chemicals, but also for the development of QSTR models based on mixture toxicity data. Studies on all these topics would allow ensuring a better estimation of the effects of mixtures, transformation products and metabolites in various environmental compartments.

REFERENCES

Aalizadeh, R., von der Ohe P. C., Thomaidis, N. S., 2017. Prediction of acute toxicity of emerging contaminants on the water flea *Daphnia magna* by ant colony optimization-support vector machine QSTR models. Environmental Science: Processes and Impacts, 19, 438-448.

Abooali, D., Sobati, M. A., 2014. Novel method for prediction of normal boiling point and enthalpy of vaporization at normal boiling point of pure refrigerants: A QSPR approach. International Journal of Refrigeration, 40, 282-293.

Aruoja, V., Moosus, M., Kahru, A., Sihtmäe, M., Maran, U., 2014. Measurement of baseline toxicity and QSAR analysis of 50 non-polar and 58 polar narcotic chemicals for the alga *Pseudokirchneriella subcapitata*. Chemosphere, 96, 23-32.

Aruoja, V., Sihtmäe, M., Dubourguier, H. C., Kahru, A., 2011. Toxicity of 58 substituted anilines and phenols to algae *Pseudokirchneriella subcapitata* and bacteria *Vibrio fischeri*: Comparison with published data and QSARs. Chemosphere, 84, 1310-1320.

Bagheri, M., Borhani, T. N. G., Zahedi, G., 2012. Estimation of flash point and autoignition temperature of organic sulfur chemicals. Energy Conversion and Management, 58, 185-196.

Basant, N., Gupta, S., Singh, K. P., 2015. Predicting aquatic toxicities of chemical pesticides in multiple test species using nonlinear QSTR modeling approaches. Chemosphere, 139, 246-255.

Becke, A. D., 1993. Density-functional thermochemistry. III. The role of exact exchange. Journal of Chemical Physics, 98, 5648-5652.

Bhhatarai, B., Gramatica, P., 2010a. Oral LD_{50} toxicity modeling and prediction of per- and polyfluorinated chemicals on rat and mouse. Molecular Diversity, 15, 467-476.

Bhhatarai, B., Gramatica, P., 2010b. Per- and polyfluoro toxicity (LC_{50} inhalation) study in rat and mouse using QSAR modeling. Chemical Research in Toxicology, 23, 528-539.

Bhhatarai, B., Gramatica, P., 2011a. Modelling physico-chemical properties of (benzo)triazoles, and screening for environmental partitioning. Water Research, 45, 1463-1471.

Bhhatarai, B., Gramatica, P., 2011b. Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. Environmental Science and Technology, 45, 8120-8128.

Boyce, D. G., Lewis, M. R., Worm, B., 2010. Global phytoplankton decline over the past century. Nature, 466, 591-596.

Brinkmann, M., Blenkle, H., Salowsky, H., Bluhm, K., Schiwy, S., Tiehm, A., Hollert, H., 2014. Genotoxicity of heterocyclic PAHs in the micronucleus assay with the fish liver cell line RTL-W1. PLoS ONE, 9, 1-8.

Brown, A. R., Whale, G., Jackson, M., Marshall, S., Hamer, M., Solga, A., Kabouw, P., Galay-Burgos, M., Woods, R., Nadzialek, S., Maltby, L., 2016. Toward the definition of specific protection goals for the environmental risk assessment of chemicals: A perspective on environmental regulation in Europe. Integrated Environmental Assessment and Management, 13, 17-37.

Buttarelli, F. R., Pellicano, C., Pontieri, F. E., 2008. Neuropharmacology and behavior in planarians: Translations to mammals. Comparative Biochemistry and Physiology, Part C, 147, 399-408.

CAS Registry. http://support.cas.org/content/chemical-substances. (accessed Apr 2018).

Cassani, S., Kovarich, S., Papa, E., Roy, P. P., van der Wal, L., Gramatica, P., 2013. Daphnia and fish toxicity of (benzo)triazoles: Validated QSAR models, and interspecies quantitative activity-activity modelling. Journal of Hazardous Materials, 258-259, 50-60.

Cassotti, M., Ballabio, D., Todeschini, R., Consonni, V., 2015. A similarity-based QSAR model for predicting acute toxicity towards the fathead minnow (*Pimephales promelas*). SAR and QSAR in Environmental Research, 26, 217-243.

Castaño, A., Bols, N., Braunbeck, T., Dierickx, P., Halder, M., Isomaa, B., Kawahara, K., Lee, L. E. J., Mothersill, C., Pärt, P., Repetto, G., Sintes, J. R., Rufli, H., Smith, R., Wood, C., Segner, H., 2003. The use of fish cells in ecotoxicology. Alternatives to Laboratory Animals, 31, 317-351.

Chen, C. Y., Kuo, K. L., Fan, J. W., 2012. Toxicity of propargylic alcohols on green alga-*Pseudokirchneriella subcapitata*. Journal of Environmental Monitoring, 14, 181-186.

Chen, L. Chen, X., 2012. Results of molecular docking as descriptors to predict human serum albumin binding affinity. Journal of Molecular Graphics and Modelling, 33, 35-43.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L. Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR modeling: Where have you been? Where are you going to? Journal of Medicinal Chemistry, 57, 4977-5010.

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.

Colmenarejo, G., Alvarez-Pedraglio, A., Lavandera, J. L., 2001. Cheminformatic models to predict binding affinities to human serum albumin. Journal of Medicinal Chemistry, 44, 4370-4378.

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.

Cronin M. T. D., Cronin M. T. D., Madden, J. C. (Eds), 2010. Biological read-across: Mechanistically-based species-species and endpoint-endpoint extrapolations. *In Silico* Toxicology: Principles and Applications. RSC Publishing, Cambridge, UK.

Cronin, M. T. D., 2017. (Q)SARs to predict environmental toxicities: Current status and future needs. Environmental Science: Processes and Impacts, 19, 213-220.

Danish(Q)SAR Database, Division of Diet, Disease Prevention and Toxicology, National Food Institute, Technical University of Denmark. http://qsar.food.dtu.dk (accessed July 2018).

Devinyak, O., Havrylyuk , D., Lesyk R., 2014. 3D-MoRSE descriptors explained. Journal of Molecular Graphics and Modelling, 54, 194-203.

DRAGON v. 6, 2013. Talete srl, Milan, Italy. http://www.talete.mi.it/ (accessed May 2018).

Dulin, F., Halm-Lemeille, M. P., Lozano, S., Lepailleur, A., Santos, J. S., Rault, S., Bureau, R., 2012. Interpretation of honeybees contact toxicity associated to acetylcholinesterase inhibitors. Ecotoxicology and Environmental Safety, 79, 13-21.

Dyer, S. D., Versteeg, D. J., Belanger, S. E., Chaney, J. G., Mayer, F. L., 2006. Interspecies correlation estimates predict protective environmental concentrations. Environmental Science and Technology, 40, 3102-3111.

EC, 2006. European Commission, Regulation No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Official Journal of the European Union, L 396/1-849.

ECHA Registered Substances Database. http://echa.europa.eu/en/information-on-chemicals/registered-substances (accessed July 2018).

Enoch, S. J., Puzyn, T, Leszczynski, J., Cronin, M. T. D. (Eds), 2010. The use of quantum mechanics derived descriptors in computational toxicology. Recent Advances in QSAR Studies. Methods and Applications. Springer, New York, USA.

Eroglu, E., Palaz, S., Oltulu, O., Turkmen, H., Ozaydın, C., 2007. Comparative QSTR study using semiempirical and first principle methods based descriptors for acute toxicity of diverse organic compounds to the fathead minnow. International Journal of Molecular Sciences, 8,1265-1283.

Ertürk, M. D., Saçan, M. T., 2012. First toxicity data of chlorophenols on marine alga *Dunaliella tertiolecta*: Correlation of marine algal toxicity with hydrophobicity and interspecies toxicity relationships. Environmental Toxicology and Chemistry, 31, 1113-1120.

Ertürk, M. D., Saçan, M. T., 2013. Assessment and modeling of the novel toxicity data set of phenols to *Chlorella vulgaris*. Ecotoxicology and Environmental Safety, 90, 61-68.

Ertürk, M. D., Saçan, M. T., Novič, M., Minovksi, N., 2012. Quantitative structure-activity relationships (QSARs) using the novel marine algal toxicity data of phenols. Journal of Molecular Graphics and Modelling, 38, 90-100.

Ferguson, J., 1939. The use of chemical potentials as indices of toxicity. Proceedings of the Royal Society B: Biological Sciences, 127, 387-404.

Fjodorova, N., Vračko, M., Novič, M., Roncaglioni, A., Benfenati, E., 2010. New public QSAR model for carcinogenicity. Chemistry Central Journal, 4 (Suppl 1), S3.

Free, S. M., Wilson, J. W., 1964. A mathematical contribution to structure-activity studies. Journal of Medicinal Chemistry, 7, 395-399.

Fu, L., Li, J. J., Wang, Y., Wang, X. H., Wen, Y., Qin, W. C., Su, L. M., Zhao, Y. H., 2015. Evaluation of toxicity data to green algae and relationship with hydrophobicity. Chemosphere, 120, 16-22.

Furuhama, A., Hasunuma, K., Aoki, Y., 2015. Interspecies quantitative structure-activity-activity relationships (QSAARs) for prediction of acute aquatic toxicity of aromatic amines and phenols. SAR and QSAR in Environmental Research, 26, 301-323.

Golbraikh, A., Muratov, E., Fourches, D., Tropsha, A., 2014. Data set modelability by QSAR. Journal of Chemical Information and Modeling, 54, 1-4.

Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y., Lee, K., Tropsha, A., 2003. Rational selection of training and test sets for the development of validated QSAR models. Journal of Computer-Aided Molecular Design, 17, 241-253.

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

Gramatica, P. 2007. Principles of QSAR models validation: Internal and external. QSAR and Combinatorial Science, 26, 694-701.

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.

Gramatica, P., Cassani, S., Roy, P. P., Kovarich, S., Yap, C. W., Papa, E., 2012. QSAR modeling is not "push a button and find a correlation": A case study of toxicity of (benzo-)triazoles on algae. Molecular Informatics, 31, 817-835.

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., Giani, E., Papa, E., 2007a. Statistical external validation and consensus modeling: A QSPR case study for K_{oc} prediction. Journal of Molecular Graphics and Modelling, 25, 755-766.

Gramatica, P., Papa, E., 2005. An update of the BCF QSAR model based on theoretical molecular descriptors. QSAR and Combinatorial Science, 24, 935-960.

Gramatica, P., Papa, E., 2007. Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure. Environmental Science and Technology, 41, 2833-2839.

Gramatica, P., Pilutti, P., Papa, E., 2007b. Approaches for externally validated QSAR modelling of nitrated polycyclic aromatic hydrocarbon mutagenicity. SAR and QSAR in Environmental Research, 18, 169-178.

Gramatica, P., Sangion, A., 2016. A historical excursus on the statistical validation parameters for QSAR models: A clarification concerning metrics and terminology. Journal of Chemical Information and Modeling, 56, 1127-1131.

Hagstrom, D., Cochet-Escartin, O., Zhang, S., Khuu, C., Collins, E. M. S., 2015. Freshwater planarians as an alternative animal model for neurotoxicology. Toxicological Sciences, 147, 270-285.

Hansch, C., Fujita, T., 1964. ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. Journal of American Chemical Society, 86, 1616-1626.

Hansch, C., Glave, W. R., 1971. Structure-activity relationships in membrane-perturbing agents. Hemolytic, narcotic, and antibacterial compounds. Molecular Pharmacology, 7, 337-354.

Hansch, C., Maloney, P. P., Fujita, T., Muir, R. M., 1962. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature, 194, 178-180.

Helfman, G. S., Collette, B. B., Facey, D. E., Bowen, B. W., 2009. The Diversity of Fishes: Biology, Evolution, and Ecology, Wiley-Blackwell, Hoboken, NJ, USA.

Jackson, J. E., 1991. A User's Guide to Principal Components. Wiley and Sons, Inc., New York, USA.

Joanes, D. N., Gill, C. A., 1998. Comparing measures of sample skewness and kurtosis. Statistician, 47, 183-189.

Judson, R., Richard, A., Dix, D. J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T., Smith, E., 2009. The toxicity data landscape for environmental chemicals. Environmental Health Perspectives, 117, 685–695.

Kar, S., Das, R. N., Roy, K., Leszczynski, J., 2016. Can toxicity for different species be correlated? The concept and emerging applications of interspecies quantitative structure-toxicity relationship (i-QSTR) modeling. International Journal of Quantitative Structure-Property Relationships, 1, 23-51.

Kar, S., Deeb, O., Roy, K., 2012. Development of classification and regression based QSAR models to predict rodent carcinogenic potency using oral slope factor. Ecotoxicology and Environmental Safety, 82, 85-95.

Kar, S., Roy, K., 2010. First report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. Chemosphere, 81, 738-747.

Kawakatsu, M., Oki, I., Tamura, S., 1995. Taxonomy and geographical distribution of *Dugesia japonica* and *D. ryukyuensis* in the Far East. Hydrobiologia, 305, 55-61.

Khajeh, A., Modarress, H., 2011. Quantitative structure-property relationship for flash points of alcohols. Industrial and Engineering Chemistry Research, 50, 11337-11342.

Könemann, H., 1981. Quantitative structure-activity relationships in fish toxicity studies. Part I. Relationship for 50 industrial pollutants. Toxicology, 19, 209-221.

Lee, C., Yang, W., Parr , R. G., 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B, 37, 785-789.

Lessigiarska, I., Worth, A. P., Sokull-Klüttgen, B., Jeram, S., Dearden, J. C., Netzeva, T. I., Cronin, M. T. D., 2004. QSAR investigation of a large data set for fish, algae and *Daphnia* toxicity. SAR and QSAR in Environmental Research, 15, 413-431.

Li, F., Wu, H., Li, L., Li, X., Zhao, J., Peijnenburg, W., 2012. Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. Ecotoxicology and Environmental Safety, 80, 273-279.

Li, M. H., 2008. Effects of nonionic and ionic surfactants on survival, oxidative stress, and cholinesterase activity of planarian. Chemosphere, 70, 1796-1803.

Li, M. H., 2012a. Acute toxicity of benzophenone-type UV filters and paraben preservatives to freshwater planarian, *Dugesia japonica*. Toxicological and Environmental Chemistry, 94, 566-573.

Li, M. H., 2012b. Survival, mobility, and membrane-bound enzyme activities of freshwater planarian, *Dugesia japonica*, exposed to synthetic and natural surfactants. Environmental Toxicology and Chemistry, 31, 843-850.

Li, M. H., 2013a. Acute toxicity of industrial endocrine-disrupting chemicals, natural and synthetic sex hormones to the freshwater planarian, *Dugesia japonica*. Toxicological and Environmental Chemistry, 95, 984-991.

Li, M. H., 2013b. Acute toxicity of 30 pharmaceutically active compounds to freshwater planarians, *Dugesia japonica*. Toxicological and Environmental Chemistry, 95, 1157-1170.

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.

Liu, H., Papa, E., Gramatica, P., 2006. QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. Chemical Research in Toxicology, 19, 1540-1548.

Liu, H., Sun, P., Liu, H., Yang, S., Wang, L., Wang, Z., 2015. Acute toxicity of benzophenone-type UV filters for *Photobacterium phosphoreum* and *Daphnia magna*: QSAR analysis, interspecies relationship and integrated assessment. Chemosphere, 135, 182-188.

Liu, H., Yang, X., Lu, R., 2016. Development of classification model and QSAR model for predicting binding affinity of endocrine disrupting chemicals to human sex hormone-binding globulin. Chemosphere, 156, 1-7.

Maggiora, G. M., 2006. On outliers and activity cliffs-why QSAR often disappoints. Journal of Chemical Information and Modeling, 46, 1535.

Masand, V. H., Mahajan, D. T., Patil, K. N., Hadda, T. B., Youssoufi, M. H., Jawarkar, R. D., Shibi, I. G., 2013. Optimization of antimalarial activity of synthetic prodiginines: QSAR, GUSAR, and CoMFA analyses. Chemical Biology and Drug Design, 81, 527-536.

McFarland, J. W., 1970. On the parabolic relationship between drug potency and hydrophobicity. Journal of Medicinal Chemistry, 13, 1092-1196.

Meyer, H., 1899. Zur theorie der alkohol narkose. Archiv für Experimentelle Pathologie Pharmakologie, 42, 109-118.

Minovski, N., Župerl, Š., Drgan, V., Novič, M., 2013. Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study. Analytica Chimica Acta, 759, 28-42.

Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., Matsushita, Y., 1992. Simple method of calculating octanol/water partition coefficient. Chemical and Pharmaceutical Bulletin, 40, 127-130.

OECD, 1987. Decision-Recommendation of the Council on the Systematic Investigation of Existing Chemicals [C(87)90/Final]. Paris, France.

OECD, 1991. Decision-Recommendation of the Council on the Co-operative Investigation and Risk Reduction of Existing Chemicals [C(90)163/Final]. Paris, France.

OECD, 1992. OECD Guidelines for the Testing of Chemicals. Test No. 203, Fish Acute Toxicity Test. Paris, France.

OECD, 2004. OECD Guidelines for the Testing of Chemicals. Test No. 202, *Daphnia* sp., Acute Immobilisation Test. Paris, France.

OECD, 2006. OECD Guidelines for the Testing of Chemicals. Test No. 201, Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Paris, France.

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, 2009. The 2007 OECD List of High Production Volume Chemicals. Environment, Health and Safety Publications. Series on Testing and Assessment No. 112. Paris, France.

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

Oliveira, K. M. G., Takahata, Y., 2008. QSAR modeling of nucleosides against amastigotes of leishmania donovani using logistic regression and classification tree. QSAR and Combinatorial Science, 27, 1020-1027.

Overton, E., 1899. Ueber die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutliche Ursachen und ihre Bedeutung für die Physiologie. Vierteljahres sehr. Naturforsch Ges, 44, 88-114.

Önlü, S., Saçan, M. T., 2017a. Impact of geometry optimization methods on QSAR modelling: A case study for predicting human serum albumin binding affinity. SAR and QSAR in Environmental Research, 28:6, 491-509.

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

Önlü, S., Saçan, M. T., 2017c. An *in silico* approach to cytotoxicity of pharmaceuticals and personal care products on the rainbow trout liver cell line RTL-W1. Environmental Toxicology and Chemistry, 36, 1162-1169.

Önlü, S., Saçan, M. T., 2018. Toxicity of contaminants of emerging concern to *Dugesia japonica*: QSTR modeling and toxicity relationship with *Daphnia magna*. Journal of Hazardous Materials, 351, 20-28.

Pan, Y., Zhang, Y., Jiang, J., Ding, L., 2014. Prediction of the self-accelerating decomposition temperature of organic peroxides using the quantitative structure-property relationship (QSPR) approach. Journal of Loss Prevention in the Process Industries, 31, 41-49.

Pandith, A. H., Islam, N., 2013. Comparative assessment of QSTR models based on density functional, Hartree-Fock, AM1, and PM3 methods for acute toxicity of aliphatic compounds toward *Vibrio fischeri*. International Journal of Quantum Chemistry, 113, 830-839.

Papa, E., Gramatica, P., 2008. Externally validated QSPR modelling of VOC tropospheric oxidation by NO_3 radicals. SAR and QSAR in Environmental Research, 19, 655-668.

Papa, E., Gramatica, P., 2010. QSPR as a support for the EU REACH regulation and rational design of environmentally safer chemicals: PBT identification from molecular structure. Green Chemistry, 12, 836-843.

Papa, E., Kovarich, S., Gramatica, P., 2009. Development, validation and inspection of the applicability domain of QSPR models for physicochemical properties of polybrominated diphenyl ethers. QSAR and Combinatorial Science, 28, 790-796.

Papa, E., Pilutti, P., Gramatica, P., 2008. Prediction of PAH mutagenicity in human cells by QSAR classification. SAR and QSAR in Environmental Research, 19, 115-127.

Papa, E., Villa, F., Gramatica, P., 2005. Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in *Pimephales promelas* (fathead minnow). Journal of Chemical Information and Modeling, 45, 1256-1266.

Pasha, F. A., Srivastava, H. K., Singh, P. P., 2005. Comparative QSAR study of phenol derivatives with the help of density functional theory. Bioorganic and Medicinal Chemistry, 13, 6823-6829.

Pasha, F. A., Srivastava, H. K., Srivastava, A., Singh, P. P., 2007. QSTR study of small organic molecules against *Tetrahymena pyriformis*. QSAR and Combinatorial Science, 26, 69-84.

Petersson, G. A., Bennett, A., Tensfeldt, T. G., Al-Laham, M. A., Shirley, W.A., Mantzaris, J., 1988. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row atoms. Journal of Chemical Physics, 89, 2193-2218.

Petrie, B., Barden, R., Kasprzyk-Hordern, B., 2015. A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring. Water Research, 72, 3-27.

Pirsig, R. M., 1974. Zen and the Art of Motorcycle Maintenance: An Inquiry into Values. William Morrow and Company, New York, USA.

Pramanik, S., Roy, K., 2014. Predictive modeling of chemical toxicity towards *Pseudokirchneriella subcapitata* using regression and classification based approaches. Ecotoxicology and Environmental Safety, 101, 184-190.

Puzyn, T., Suzuki, N., Haranczyk, M., Rak, J., 2008. Calculation of quantum-mechanical descriptors for QSPR at the DFT level: Is it necessary? Journal of Chemical Information and Modeling, 48, 1174-1180.

R Development Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (accessed July 2018).

Raimondo, S., Lilavois, C. R., Barron, M. G., 2015. Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity: User Manual, Version 3.3, EPA/600/R-15/192, US Environmental Protection Agency, Office of Research and Development, Gulf Ecology Division, Gulf Breeze, FL, USA.

Reenu, Vikas, 2015. Exploring the role of quantum chemical descriptors in modeling acute toxicity of diverse chemicals to *Daphnia magna*. Journal of Molecular Graphics and Modelling, 61, 89-101.

Richet, C., 1893. Sur le rapport entre la toxicité et les propriétés physiques des corps. CR Soc Biol, 54, 775-776.

Rinnan, A., Christensen, N. J., Engelsen, S. B., 2010. How the energy evaluation method used in the geometry optimization step affect the quality of the subsequent QSAR/QSPR models. Journal of Computer Aided Molecular Design, 24, 17-22.

Roy, K., Das, R. N., 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.

Roy, K., Kar, S., Ambure, P., 2015b. On a simple approach for determining applicability domain of QSAR models. Chemometrics and Intelligent Laboratory Systems, 145, 22-29.

Roy, K., Kar, S., Das, R. N., 2015a. A Primer on QSAR/QSPR Modeling: Fundamental Concepts. Springer, London, UK.

Roy, P. P., Kovarich, S., Gramatica, P., 2011. QSAR model reproducibility and applicability: A case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles. Journal of Computational Chemistry, 32, 2386-2396.

Russel, W. M. S., Burch, R. L., 1959. The Principles of Humane Experimental Technique. Methuen & Co. Ltd., London, UK. http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc (accessed May 2018).

Ruusmann ,V., Sild, S., Maran, U., 2015. QSAR DataBank repository: Open and linked qualitative and quantitative structure-activity relationship models. Journal of Cheminformatics, 7, 32.

Rybinska, A., Sosnowska, A., Barycki, M., Puzyn, T., 2016. Geometry optimization method versus predictive ability in QSPR modeling for ionic liquids. Journal of Computer-Aided Molecular Design, 30, 165-176.

Sangion, A., Gramatica, P., 2016. Hazard of pharmaceuticals for aquatic environment: Prioritization by structural approaches and prediction of ecotoxicity. Environment International, 95, 131-143.

Sauvé, S., Desrosiers, M., 2014. A review of what is an emerging contaminant. Chemistry Central Journal, 8, 15-21.

Schirmer , K., 2006. Proposal to improve vertebrate cell cultures to establish them as substitutes for the regulatory testing of chemicals and effluents using fish. Toxicology, 224, 163-183.

Schlegel, H. B., 2011. Geometry optimization. WIREs Computational Molecular Science, 1, 790-809.

Schneider, G., Neidhart, W., Giller, T., Schmid, G., 1999. “Scaffold-Hopping” by topological pharmacophore search: A contribution to virtual screening. Angewandte Chemie International Edition, 38, 2894-2896.

Schnell, S., Bols, N. C., Barata, C., Porte, C., 2009. Single and combined toxicity of pharmaceuticals and personal care products (PPCPs) on the rainbow trout liver cell line RTL-W1. Aquatic Toxicology, 93, 244-252.

Schreer, A., Tinson, C., Sherry, J. P., Schirmer, K., 2005. Application of Alamar blue/5-carboxyfluorescein diacetate acetoxymethyl ester as a noninvasive cell viability assay in primary hepatocytes from rainbow trout. Analytical Biochemistry, 344, 76-85.

Schüürmann, G., Cronin, M. T. D., Livingstone, D. J. (Eds), 2004. Quantum chemical descriptors in structure-activity relationships – calculation, interpretation and comparison of methods. Predicting Chemical Toxicity and Fate. CRC Press, Boca Raton, FL, USA.

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.

Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S., Dial, S. L., Moland, C. L., Sheehan, D. M., 2001. QSAR models using a large diverse set of estrogens. Journal of Chemical Information and Computer Sciences, 41, 186-195.

Singh, K. P., Gupta, S., Kumar, A., Mohan, D., 2014. Multispecies QSAR modeling for predicting the aquatic toxicity of diverse organic chemicals for regulatory toxicology. Chemical Research in Toxicology, 27, 741-753.

Sobanska, M. A., Cesnaitis, R., Sobanski, T., Versonnen, B., Bonnomet, V., Tarazona, J. V., De Coen, W., 2014. Analysis of the ecotoxicity data submitted within the framework of the REACH Regulation. Part 1. General overview and data availability for the first registration deadline. Science of the Total Environment, 470-471, 1225-1232.

SPARTAN v. 10, 2011. Wavefunction, Inc., Irvine, CA, USA. http://wavefun.com (accessed May 2018).

Stewart, J. J. P., 2007. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. Journal of Molecular Modeling, 13, 1173-1213.

Stollewerk, A., 2010. The water flea *Daphnia* - a 'new' model system for ecology and evolution? Journal of Biology, 9, 21-24.

SU0303, 2015. TÜBİTAK-ARDEB Su-Entegre Havza Yönetimi Çağrı Programı 'SU0303 - Havza Ölçeğinde Noktasal ve Yayılı Kirletici Kaynakların Yönetimi' Çağrı Metni. https://www.tubitak.gov.tr/sites/default/files/su0303_cagri_metni.pdf (accessed May 2018)

Tanneberger, K., Knöbel, M., Busser, F. J. M., Sinnige, T. L., Hermens, J. L. M., Schirmer K., 2013. Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. Environmental Science and Technology, 47, 1110-1119.

Tarazona, J. V., Sobanska, M. A., Cesnaitis, R., Sobanski, T., Bonnomet, V., Versonnen, B., De Coen, W., 2014. Analysis of the ecotoxicity data submitted within the framework of the REACH Regulation. Part 2. Experimental aquatic toxicity assays. Science of the Total Environment, 472, 137-145.

TC, 2017. Çevre ve Şehircilik Bakanlığı, Kimyasalların Kaydı, Değerlendirilmesi, İzni ve Kısıtlanması Hakkında Yönetmelik (KKDİK). Türkiye Cumhuriyeti Resmi Gazete, 23 Haziran 2017, Sayı: 30105 (Mükerrer), 464-855.

Tee, P. N. C., Wong, Y. T. J., Sherry, J. P., Bols, N. C., 2011. Effect of acid blue 80, an anthracenedione dye, on rainbow trout liver, gill and gut cells *in vitro*. Ecotoxicology and Environmental Safety, 74, 1874-1878.

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

Todeschini, R., Consonni, V., Maiocchi, A., 1999. The *K* correlation index: Theory development and its application in chemometrics. Chemometrics and Intelligent Laboratory Systems, 46, 13-29.

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., 2010. Best practices for QSAR model development, validation, and exploitation. Molecular Informatics, 29, 476-488.

Tugcu, G., Ertürk, M. D., Saçan, M. T., 2017. On the aquatic toxicity of substituted phenols to *Chlorella vulgaris*: QSTR with an extended novel data set and interspecies models. Journal of Hazardous Materials, 339, 122-130.

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.

US EPA, 2016. The United States Environmental Protection Agency Web-ICE v3.3. https://www3.epa.gov/webice/ (accessed July 2018).

Verhaar, H. J. M., van Leeuwen, C. J., Hermens, J. L. M., 1992. Classifying environmental pollutants. 1: Structure-activity relationships for prediction of aquatic toxicity. Chemosphere, 25, 471-491.

Villain, J., Lozano, S., Halm-Lemeille, M. P., Durrieu, G., Bureau, R., 2014. Quantile regression model for a diverse set of chemicals: Application to acute toxicity for green algae. Journal of Molecular Modeling, 20, 2508.

Wang, Y., Chen, J., Yang, X., Lyakurwa, F., Li, X., Qiao, X., 2015. *In silico* model for predicting soil organic carbon normalized sorption coefficient (K_{OC}) of organic chemicals. Chemosphere, 119, 438-444.

Wu, X., Zhang, Q., Hu, J., 2016. QSAR study of the acute toxicity to fathead minnow based on a large dataset. SAR and QSAR in Environmental Research, 27, 147-164.

Zupan, J., Gasteiger, J., 1999. Neural Networks in Chemistry and Drug Design. 2nd edition, Wiley-VCH, Weinheim, Germany, 380.

Zupan, J., Novič, M., Ruisánchez, I., 1997. Kohonen and counterpropagation artificial neural networks in analytical chemistry. Chemometrics and Intelligent Laboratory Systems, 38, 1-23.

APPENDIX A: DETAILED RESULTS OF CASE STUDY MODELS

Table A.1. Case study dataset chemicals, experimental/predicted log K_{HSA}, model descriptors.

ID	CAS	Name	Exp log K_{HSA}	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
1	50-78-2	Acetylsalicylic acid	-1.39	-0.82	0	1	0	1.202
2	55268-75-2	Cefuroxime*	-1.33	-0.92	0	1	1	-0.135
3	26787-78-0	Amoxicillin	-1.21	-1.13	0	1	0	0.210
4	15686-71-2	Cephalexin	-1.11	-1.07	0	1	0	0.412
5	2022-85-7	5-fluorocytosine	-1.11	-1.24	0	0	0	-1.079
6	16110-51-3	Cromolyn	-1.07	-0.55	2	2	0	1.720
7	58-08-2	Caffeine	-0.92	-0.94	0	0	0	-0.100
8	103-90-2	Acetaminophen	-0.81	-0.69	0	0	0	0.683
9	73-22-3	L-tryptophan*	-0.78	-0.81	0	1	0	1.249
10	59-05-2	Methotrexate	-0.77	-0.85	1	2	1	0.384
11	51-52-5	Propylthiouracil	-0.75	-0.48	0	0	0	1.381
12	60-80-0	Antipyrine	-0.69	-0.40	0	0	0	1.620
13	87-08-1	Phenoxymethylpenicillinic acid	-0.69	-0.46	0	1	1	1.326
14	69-72-7	Salicylic acid	-0.66	-0.83	0	1	0	1.167
15	64544-07-6	Cefuroxime axetil	-0.56	-0.63	0	0	1	-0.147
16	33419-42-0	Etoposide*	-0.49	-0.22	2	0	0	0.910
17	29122-68-7	Atenolol	-0.48	-0.38	0	0	1	0.669
18	56-75-7	Chloramphenicol	-0.46	-0.59	0	0	0	1.020
19	51481-61-9	Cimetidine	-0.44	-0.40	0	0	1	0.610
20	94-20-2	Chlorpropamide	-0.44	-0.27	0	0	0	2.058
21	3930-20-9	Sotalol	-0.44	-0.60	0	0	0	0.967
22	58-93-5	Hydrochlorothiazide	-0.42	-0.69	1	0	0	0.038
23	1156-19-0	Tolazamide*	-0.42	-0.13	0	0	0	2.496
24	50-23-7	Hydrocortisone	-0.40	-0.13	2	0	0	1.217
25	42200-33-9	Nadolol	-0.40	-0.35	1	0	0	1.146
26	50-24-8	Prednisolone	-0.40	-0.14	2	0	0	1.194
27	51-34-3	Scopolamine	-0.34	-0.33	0	0	1	0.824
28	26839-75-8	Timolol	-0.33	-0.56	0	0	0	1.128
29	37350-58-6	Metoprolol	-0.29	-0.04	0	0	1	1.757
30	738-70-5	Trimethoprim*	-0.26	-0.11	0	0	1	1.544
31	1091-85-6	Dansylglycine	-0.26	-0.48	1	1	0	1.676
32	137-58-6	Lidocaine	-0.23	-0.03	0	0	0	2.840

Table A.1. Continued.

ID	CAS	Name	Exp log K_{HSA}	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
33	83-43-2	Methylprednisolone	-0.22	-0.06	2	0	0	1.446
34	64-77-7	Tolbutamide	-0.22	-0.18	0	0	0	2.336
35	526-08-9	Sulfaphenazole	-0.21	0.16	0	0	1	2.395
36	37517-30-9	Acebutolol	-0.21	-0.08	0	0	1	1.615
37	59-46-1	Procaine*	-0.19	-0.04	0	0	1	1.775
38	63590-64-7	Terazosin	-0.16	0.08	1	0	1	1.488
39	6452-71-7	Oxprenolol	-0.15	0.11	0	0	1	2.232
40	84057-84-1	Lamotrigine	-0.13	-0.15	0	0	0	2.427
41	4205-90-7	Clonidine	-0.13	0.03	0	0	0	3.021
42	13523-86-9	Pindolol	-0.13	0.01	0	0	1	1.926
43	54-31-9	Furosemide	-0.13	-0.34	0	1	1	1.714
44	298-46-4	Carbamazepine*	-0.10	-0.08	0	0	0	2.679
45	66357-35-5	Ranitidine	-0.10	-0.13	0	0	1	1.466
46	7689-03-4	Camptothecin	-0.08	0.04	2	0	0	1.746
47	60-54-8	Tetracycline	-0.08	-0.33	3	0	1	-1.111
48	34841-39-9	Bupropion	-0.05	0.09	0	0	0	3.227
49	103628-46-2	Sumatriptan	-0.05	-0.08	0	0	1	1.631
50	81-81-2	Warfarin	-0.04	0.12	1	0	0	2.662
51	28395-03-1	Bumetanide*	-0.03	-0.28	0	1	0	2.960
52	129-20-4	Oxyphenbutazone	-0.02	0.24	0	0	0	3.687
53	87848-99-5	Acrivastine	-0.02	0.22	0	1	0	4.576
54	57-41-0	Phenytoin	0.00	-0.25	0	0	0	2.105
55	564-25-0	Doxycycline	0.01	-0.29	3	0	1	-0.967
56	22071-15-4	Ketoprofen	0.03	-0.16	0	1	0	3.336
57	13655-52-2	Alprenolol	0.04	-0.09	0	0	0	2.640
58	19216-56-9	Prazosin*	0.06	0.27	1	0	1	2.108
59	100986-85-4	Levofloxacin	0.14	0.23	3	1	1	1.636
60	85721-33-1	Ciprofloxacin	0.14	-0.24	1	1	1	1.410
61	36894-69-6	Labetalol	0.14	0.14	0	0	1	2.330
62	70458-96-7	Norfloxacin	0.14	-0.28	1	1	1	1.269
63	50-33-9	Phenylbutazone	0.19	0.32	0	0	0	3.954
64	22204-53-1	Naproxen	0.25	-0.12	1	1	0	2.824
65	637-07-0	Clofibrate*	0.27	0.13	0	0	0	3.330
66	525-66-6	Propranolol	0.28	0.08	1	0	0	2.540
67	94-24-6	Tetracaine	0.32	0.33	0	0	1	2.962
68	6990 06 3	Fusidic acid	0.33	0.77	2	1	0	5.072
69	303-81-1	Novobiocin	0.35	0.67	1	0	1	3.400

Table A.1. Continued.

ID	CAS	Name	Exp log K_{HSA}	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
70	99614-02-5	Ondansetron	0.37	0.23	0	0	1	2.635
71	548-73-2	Droperidol	0.43	0.36	0	0	1	3.049
72	56-54-2	Quinidine*	0.44	0.46	1	0	1	2.734
73	53-86-1	Indomethacin	0.47	0.11	0	1	0	4.211
74	130-95-0	Quinine	0.49	0.46	1	0	1	2.734
75	599-79-1	Sulfasalazine	0.56	0.33	0	1	1	3.885
76	57-83-0	Progesterone	0.59	0.60	2	0	0	3.580
77	50-47-5	Desipramine	0.61	0.29	0	0	0	3.852
78	50-28-2	Estradiol	0.68	0.67	2	0	0	3.813
79	10238-21-8	Glibenclamide*	0.68	0.70	0	0	1	4.140
80	58-22-0	Testosterone	0.74	0.53	2	0	0	3.333
81	50-49-7	Imipramine	0.75	0.45	0	0	0	4.388
82	65277-42-1	Ketoconazole	0.84	0.53	0	0	1	3.610
83	58-40-2	Promazine	0.92	0.76	2	0	0	4.076
84	84625-61-6	Itraconazole	1.04	1.42	0	0	1	6.467
85	146-54-3	Triflupromazine	1.05	1.05	2	0	0	5.018
86	50-53-3	Chlorpromazine*	1.10	0.96	2	0	0	4.740
87	91161-71-6	Terbinafine	1.17	0.95	1	0	0	5.336
88	23593-75-1	Clotrimazole	1.34	0.71	0	0	0	5.223

*Test set compound.

Table A.2. Case study external set chemicals, predicted log K_{HSA}, model descriptors.

ID	CAS	Name	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
89	107-98-2	1-methoxy-2-propanol	-0.94	0	0	0	-0.111
90	120-78-5	2,2'-dithiobisbenzothiazole	0.89	0	0	0	5.790
91	39263-32-6	2-amino-5-bromobenzonitrile	-0.37	0	0	0	1.711
92	696-23-1	2-methyl-4(5)-nitroimidazole	-0.78	0	0	0	0.412
93	371-40-4	4-fluoroaniline	-0.51	0	0	0	1.289
94	107-02-8	Acroleine	-0.74	0	0	0	0.515
95	59277-89-3	Acyclovir	-1.35	0	0	0	-1.454
96	5329-14-6	Amidosulfonic acid	-1.94	0	2	0	-1.481
97	50-48-6	Amitriptyline	0.57	0	0	0	4.772
98	57-43-2	Amobarbital	-0.31	0	0	0	1.909
99	69-53-4	Ampicillin	-1.05	0	1	0	0.477
100	62-53-3	Aniline	-0.57	0	0	0	1.083
101	100-66-3	Anisol	-0.34	0	0	0	1.814
102	27589-33-9	Azosemide	0.05	0	0	1	2.047
103	100-52-7	Benzaldehyde	-0.41	0	0	0	1.589
104	65-85-0	Benzoic acid	-0.75	0	1	0	1.434
105	98-88-4	Benzoyl chloride	-0.27	0	0	0	2.036
106	100-46-9	Benzylamine	-0.61	0	0	0	0.936
107	501-53-1	Benzyl chloroformiate	-0.10	0	0	0	2.599
108	154361-50-9	Capecitabine	-0.33	0	0	1	0.820
109	66-25-1	Capronaldehyde	-0.33	0	0	0	1.853
110	83881-51-0	Cetirizine	-0.06	0	1	0	3.676
111	367-21-5	Chlorfluoroaniline	-0.30	0	0	0	1.953
112	57-62-5	Chlortetracycline	-0.17	3	0	1	-0.596
113	882-09-7	Clofibric acid	-0.35	0	1	0	2.730
114	1622-61-3	Clonazepam	0.05	0	0	0	3.073
115	5251-34-3	Cloprednol	-0.16	2	0	0	1.118
116	486-56-6	Cotinine	-0.81	0	0	0	0.297
117	439-14-5	Diazepam	0.14	0	0	0	3.385
118	15307-86-5	Diclofenac	0.15	0	1	0	4.348
119	111-42-2	Diethanolamine	-1.30	0	0	0	-1.295
120	42399-41-7	Diltiazem	0.41	0	0	1	3.200
121	58-73-1	Diphenhydramine	0.14	0	0	0	3.385
122	46755-94-6	Diphenylpropanediol	-0.12	0	0	0	2.532
123	1141-88-4	Dithiodianiline	0.04	0	0	0	3.046
124	112-54-9	Dodecanal	0.52	0	0	0	4.590
125	106-89-8	Epichlorohydrine	-0.73	0	0	0	0.562

Table A.2. Continued.

ID	CAS	Name	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
126	57-63-6	Ethinylestradiol	1.00	2	0	0	4.861
127	91-53-2	Ethoxyquin	0.29	1	0	0	3.221
128	74-96-4	Ethyl bromide	-0.49	0	0	0	1.340
129	104227-87-4	Famciclovir	-0.72	0	0	1	-0.433
130	76824-35-6	Famotidine	-0.92	0	0	1	-1.085
131	98319-26-7	Finasteride	0.80	2	0	1	3.174
132	79660-72-3	Fleroxacin	-0.04	1	1	1	2.062
133	78755-81-4	Flumazenil	-0.15	1	0	0	1.797
134	1622-62-4	Flunitrazepam	-0.03	0	0	0	2.820
135	490-79-9	Gentisic acid	-0.92	0	1	0	0.900
136	13311-84-7	Flutamide	0.00	0	0	0	2.917
137	111-30-8	Glutaraldehyde	-0.86	0	0	0	0.141
138	68-88-2	Hydroxyzine	0.17	0	0	0	3.467
139	15687-27-1	Ibuprofen	-0.09	0	1	0	3.582
140	54-85-3	Isoniazid	-1.15	0	0	0	-0.810
141	4759-48-2	Isotretinoin	0.52	0	1	0	5.526
142	108-31-6	Maleic anhydride	-0.83	0	0	0	0.253
143	105-53-3	Malonic acid diethylester	-0.67	0	0	0	0.755
144	2898 12 6	Medazepam	0.38	0	0	0	4.140
145	61-68-7	Mefenamic acid	0.03	0	1	0	3.957
146	443-48-1	Metronidazole	-1.01	0	0	0	-0.337
147	59467-70-8	Midazolam (base)	0.64	1	0	0	4.340
148	2211-28-1	Monobenzoate	0.65	1	0	0	4.389
149	110-91-8	Morpholine	-1.07	0	0	0	-0.528
150	121-69-7	N,N'-dimethylaniline	-0.29	0	0	0	1.992
151	91-66-7	N,N-diethylaniline	-0.07	0	0	0	2.690
152	54-11-5	Nicotine	-0.52	0	0	0	1.243
153	67-20-9	Nitrofurantion	-0.81	0	0	0	0.319
154	59-87-0	Nitrofurazone	-0.84	0	0	0	0.223
155	55-63-0	Nitroglycerin	-0.83	0	0	0	0.232
156	57849-23-7	Octabase H	0.45	1	0	0	3.742
157	82419-36-1	Ofloxacin	0.23	3	1	1	1.636
158	149-73-5	Orthoformic acid trimethylester	-0.73	0	0	0	0.563
159	61869-08-7	Paroxetine	0.10	0	0	0	3.230
160	76-74-4	Pentobarbital	-0.31	0	0	0	1.909
161	50-06-6	Phenobarbital	-0.50	0	0	0	1.321
162	110-89-4	Piperidine	-0.69	0	0	0	0.701

Table A.2. Continued.

ID	CAS	Name	Pred log K_{HSA}	nR10	CATS2D_01_AN	B10[C-N]	ALOGP
163	3282-30-2	Pivaloyl chloride	-0.31	0	0	0	1.918
164	123-38-6	Propionaldehyde	-0.75	0	0	0	0.484
165	58-14-0	Pyrimethamine	-0.05	0	0	0	2.751
166	106266-06-2	Risperidone	0.64	1	0	1	3.318
167	723-46-6	Sulfamethoxazole	-0.22	0	0	1	1.182
168	57-85-2	Testosterone propionate	0.85	2	0	0	4.379
169	119-64-2	Tetralin	0.33	1	0	0	3.343
170	58-55-9	Theophylline	-1.00	0	0	0	-0.306
171	76-75-5	Thiopental	0.08	0	0	0	3.193
172	134308-13-7	Tolcapone milled	0.05	0	0	0	3.080
173	10161-34-9	Trenbolone acetate	0.42	2	0	0	3.006
174	396-01-0	Triamterene	-0.07	1	0	1	1.017
175	3380-34-5	Triclosan	0.68	0	0	0	5.116
176	935-92-2	Trimethylquinone	-0.36	0	0	0	1.747
177	6138-79-0	Triprolidine	0.36	0	0	0	4.075

APPENDIX B: DETAILED RESULTS OF ALGAE MODELS

Table B.1. Group 3 chemicals, experimental/predicted $pEC_{50,}$ model descriptors, production volume status.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
1	110-82-7	Cyclohexane	3.64	2.36	4.01	0.373	0.365	0	0	0	0	9.760	7.508	HPV
5	109-69-3	1-chlorobutane	3.41	3.55	3.03	0.539	0.230	0	0	0	0	4.957	5.748	HPV
6	75-09-2	Dichloromethane	2.04	3.21	2.39	0.589	0.044	0	0	0	0	1.860	5.552	HPV
7	75-34-3	1,1-dichloroethane*	2.97	3.25	2.39	0.565	0.037	0	0	0	0	3.301	5.563	HPV
8	107-06-2	1,2-dichloroethane	2.57	3.16	3.03	0.518	0.032	0	0	0	0	3.301	5.457	HPV
9	78-87-5	1,2-dichloropropane	2.94	3.33	3.29	0.488	0.087	0	0	0	0	4.957	5.442	HPV
10	142-28-9	1,3-dichloropropane*	3.04	3.11	3.29	0.467	0.103	0	0	0	0	4.957	5.680	N
12	67-66-3	Trichloromethane	2.71	3.39	2.39	0.575	0.029	0	0	0	0	3.301	5.403	HPV
14	79-00-5	1,1,2-trichloroethane	3.11	3.54	3.03	0.544	0.049	0	0	0	0	4.957	5.328	HPV
15	96-18-4	1,2,3-trichloropropane	3.27	3.63	3.29	0.485	0.095	0	0	0	0	6.781	5.294	HPV
16	56-23-5	Carbon tetrachloride	4.84	3.76	4.13	0.591	0.009	0	0	0	0	4.957	5.150	HPV
18	79-34-5	1,1,2,2-tetrachloroethane	3.49	3.66	3.82	0.520	0.001	0	0	0	0	6.781	5.177	HPV
19	76-01-7	Pentachloroethane	4.38	4.03	3.82	0.529	0.045	0	0	0	0	8.746	5.101	HPV
22	109-64-8	1,3-dibromopropane	3.64	3.86	3.29	0.481	0.166	0	0	0	0	6.781	5.150	N
23	75-27-4	Bromodichloromethane	4.14	4.08	4.25	0.631	0.037	0	0	0	0	4.106	4.874	N
24	124-48-1	Dibromochloromethane	4.34	4.48	4.25	0.649	0.096	0	0	0	0	4.957	4.700	N
25	96-12-8	1,2-dibromo-3-chloropropane	3.58	4.30	3.82	0.528	0.095	0	0	0	0	8.746	4.858	N
30	542-75-6	1,3-dichloropropene*	4.72	4.19	4.03	0.532	0.129	2	0	0	0	4.332	5.168	HPV
31	760-23-6	3,4-dichlorobut-1-ene	3.33	4.07	3.33	0.544	0.116	1	0	0	0	6.047	5.250	HPV
32	79-01-6	Trichloroethylene	3.30	4.45	3.33	0.615	0.057	1	0	0	0	4.332	4.710	HPV
33	127-18-4	Tetrachloroethene*	3.88	4.45	4.25	0.597	0.033	0	0	0	0	6.047	4.483	HPV
34	78-79-5	Isoprene	3.01	4.09	3.96	0.481	0.208	1	0	0	0	3.896	4.890	HPV
35	111-78-4	1,5-cyclooctadiene	4.12	4.67	4.18	0.373	0.381	4	0	0	0	7.003	5.409	HPV
36	3048-65-5	3a,4,7,7a-tetrahydro-1H-indene	4.23	4.79	4.18	0.405	0.284	4	0	0	0	8.783	5.379	HPV
37	16219-75-3	5-ethylidene-8,9,10-trinorborn-2-ene	4.39	4.77	4.39	0.433	0.267	3	0	0	0	8.783	5.185	HPV
43	71-36-3	1-butanol	1.68	2.03	1.82	0.437	0.184	0	0	0	0	0.640	6.598	HPV
44	78-83-1	Iso-Butanol	1.64	2.04	1.82	0.424	0.191	0	0	0	0	0.640	6.549	HPV
46	75-65-0	2-methyl-2-propanol	1.66	1.82	2.52	0.385	0.217	0	0	0	0	0.640	6.727	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
47	71-41-0	1-pentanol*	2.38	2.24	1.82	0.438	0.243	0	0	0	0	1.463	6.590	HPV
48	584-02-1	3-pentanol	2.13	2.22	1.82	0.435	0.243	0	0	0	0	1.463	6.595	N
49	111-27-3	Hexanol	2.95	2.53	3.26	0.448	0.326	0	0	0	0	2.519	6.592	HPV
50	111-70-6	1-heptanol*	3.53	2.81	3.17	0.450	0.400	0	0	0	0	3.765	6.591	N
51	111-87-5	1-octanol	3.67	3.59	3.63	0.455	0.474	0	0	0	0	10.150	6.591	HPV
52	143-08-8	1-nonanol	4.82	3.94	4.99	0.456	0.547	0	0	0	0	12.271	6.591	HPV
53	112-30-1	1-decanol	5.16	4.32	4.99	0.458	0.627	0	0	0	0	14.489	6.591	HPV
54	25339-17-7	Isodecyl alcohol	4.37	3.32	4.01	0.361	0.580	0	0	0	0	8.378	6.562	HPV
56	108-93-0	Cyclohexanol	2.39	2.21	2.52	0.394	0.288	0	0	0	0	1.428	6.529	HPV
57	96-23-1	1,3-dichloro-2-propanol	2.31	2.76	2.39	0.442	0.085	0	0	0	0	1.463	5.533	N
59	80-04-6	Hydrogenatedbisphenol A	3.47	3.75	4.04	0.345	0.781	0	0	0	0	8.914	6.492	N
94	109-59-1	2-isopropoxyethanol	1.21	2.22	2.42	0.410	0.109	0	0	0	0	0.108	5.971	N
95	111-76-2	2-butoxyethanol	1.81	2.70	1.89	0.449	0.239	0	0	0	0	0.499	5.898	HPV
97	112-34-5	2-(2-Butoxyethoxy)ethanol	2.17	2.77	2.63	0.429	0.253	0	0	0	0	0.311	5.734	HPV
99	60-29-7	Diethylether*	1.51	2.30	1.82	0.449	0.102	0	0	0	0	0.640	6.096	HPV
100	142-96-1	1,1'-oxybis-butane	3.77	3.44	3.68	0.443	0.435	0	0	0	0	5.170	5.993	HPV
101	111-44-4	Bis(2-chloroethyl) ether*	2.62	3.09	1.82	0.462	0.126	0	0	0	0	2.519	5.429	HPV
102	127-90-2	2,3,3,3,2',3',3'-Octachlorodipropyl ether*	5.50	5.07	3.74	0.481	0.234	0	0	0	0	16.019	4.928	N
104	75-07-0	Acetaldehyde	3.23	3.33	3.85	0.538	0.073	1	0	0	0	0.101	5.326	HPV
107	123-15-9	2-methylvaleraldehyde	4.22	3.83	4.20	0.450	0.329	1	0	0	0	2.080	5.172	N
110	170-30-3	Crotonaldehyde	4.87	4.17	3.99	0.526	0.110	3	0	0	0	0.300	4.909	HPV
113	111-30-8	Glutaraldehyde	4.72	3.76	4.77	0.442	0.217	2	0	0	0	0.006	4.985	HPV
119	693-54-9	2-decanone	4.50	5.26	4.56	0.475	0.682	0	0	0	0	13.406	5.434	N
120	112-12-9	2-undecanone*	4.95	5.66	6.02	0.474	0.776	0	0	0	0	15.626	5.433	N
121	593-08-8	2-tridecanone	6.22	6.40	6.02	0.473	0.924	0	0	0	0	20.302	5.434	N
123	108-94-1	Cyclohexanone	1.92	3.20	3.47	0.411	0.313	0	0	0	0	1.103	5.334	HPV
124	1502-22-3	2-(1'-Cyclohexenyl)cyclohexanone	3.97	4.83	4.03	0.381	0.587	1	0	0	0	7.967	4.953	N
125	78-59-1	3,5,5-trimethyl-2-cyclohexen-1-one*	2.78	4.03	5.07	0.390	0.363	1	0	0	0	3.789	4.938	HPV
127	141-78-6	Ethylacetate	1.35	2.54	1.82	0.482	0.101	0	0	0	0	0.344	5.896	HPV
129	110-19-0	Isobutyl acetate	2.50	2.79	2.63	0.435	0.222	0	0	0	0	1.886	5.858	HPV
131	111-82-0	Methyl dodecanoate	5.83	5.85	6.02	0.483	0.817	0	0	0	0	19.687	5.843	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
132	515-84-4	Ethyl trichloroacetate	3.44	3.62	4.13	0.580	0.039	0	0	0	0	2.981	5.108	N
134	105-53-3	Diethyl malonate	2.26	2.64	1.82	0.461	0.107	0	0	0	0	0.620	5.724	HPV
137	96-33-3	Methyl acrylate	4.44	3.32	3.85	0.539	0.058	1	0	0	0	0.229	5.325	HPV
138	140-88-5	Ethylacrylate	4.64	3.39	4.31	0.509	0.110	1	0	0	0	0.789	5.297	HPV
139	141-32-2	n-butyl acrylate	4.88	3.80	3.96	0.492	0.261	1	0	0	0	2.621	5.291	HPV
140	818-61-1	2-hydroxyethyl acrylate	4.29	3.49	4.29	0.515	0.040	1	0	0	0	0.003	4.915	HPV
144	97-88-1	n-butyl methacrylate	3.79	3.73	3.75	0.473	0.331	0	0	0	0	3.813	5.320	HPV
145	688-84-6	2-ethylhexyl methacrylate	4.57	4.57	4.36	0.402	0.577	0	0	0	0	9.895	5.292	HPV
146	868-77-9	2-hydroxyethyl methacrylate	2.26	3.02	2.39	0.454	0.082	0	0	0	0	0.185	5.065	HPV
147	2867-47-2	2-(dimethylamino)ethyl methacrylate	4.24	3.74	4.27	0.415	0.166	0	0	0	0	1.247	4.301	HPV
148	13048-33-4	Hexamethylene diacrylate	5.15	4.51	5.07	0.419	0.502	2	0	0	0	4.524	5.189	HPV
149	108-05-4	Vinyl acetate	3.99	3.41	4.31	0.522	0.092	1	0	0	0	0.229	5.213	HPV
152	75-98-9	Pivalic acid	3.19	2.61	2.42	0.402	0.200	0	0	0	0	0.991	5.768	HPV
154	88-09-5	2-ethyl-butanoic acid	3.20	2.78	2.63	0.426	0.228	0	0	0	0	1.886	5.838	N
155	111-14-8	Heptanoic acid	3.34	3.49	3.23	0.499	0.430	0	0	0	0	2.981	5.895	HPV
156	124-07-2	Octanoic acid	3.57	4.24	3.63	0.496	0.520	0	0	0	0	8.833	5.882	HPV
157	334-48-5	Decanoic acid	4.16	4.26	4.36	0.362	0.532	0	0	0	0	12.907	5.794	HPV
162	335-67-1	Perfluorooctanoic acid	3.63	4.24	3.74	0.482	0.027	0	0	0	0	17.390	5.664	N
163	298-12-4	Glyoxylic acid	3.35	3.97	3.33	0.586	-0.010	1	0	0	0	1.768	4.708	HPV
164	3821-81-6	A-fluoro-β-alanine	3.13	3.32	3.13	0.475	0.045	0	0	0	1	0.221	4.887	N
167	79-10-7	Acrylic acid	4.98	3.38	4.87	0.581	0.053	1	0	0	0	0.001	5.392	HPV
168	79-41-4	Methacrylic acid	3.14	3.01	3.13	0.487	0.075	0	0	0	0	0.229	5.219	HPV
169	110-44-1	Sorbic acid	3.12	4.82	3.99	0.549	0.083	4	0	0	0	1.367	4.543	HPV
170	144-62-7	Oxalic acid	3.61	3.29	4.01	0.560	-0.053	0	0	0	0	1.756	5.066	HPV
171	124-04-9	Adipic acid*	3.39	2.65	1.82	0.427	0.258	0	0	0	0	0.189	5.879	HPV
179	108-91-8	Cyclohexylamine	3.46	2.74	2.52	0.390	0.341	0	0	0	2	1.428	6.348	HPV
182	115-70-8	2-amino-2-ethylpropanediol*	3.37	2.58	1.82	0.419	0.166	0	0	0	1	0.257	5.904	N
184	109-89-7	Diethylamine	3.13	2.40	3.05	0.425	0.173	0	0	0	0	0.640	6.032	HPV
188	111-42-2	Diethanolamine/2,2'-iminodiethanol	2.13	2.56	2.23	0.399	0.119	0	0	0	0	0.840	5.603	HPV
189	121-44-8	Triethylamine	4.10	3.00	4.01	0.395	0.252	0	0	0	0	2.519	5.556	HPV
195	102-81-8	2-(dibutylamino)ethanol*	3.92	4.12	4.12	0.414	0.558	0	0	0	2	4.055	5.541	HPV
196	124-09-4	1,6-hexanediamine	3.79	2.92	3.26	0.441	0.381	0	0	0	2	0.499	6.331	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
197	6864-37-5	2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)	4.59	5.11	4.04	0.355	0.924	0	0	0	6	8.914	6.303	HPV
200	111-18-2	N,N,N',N'-tetramethylhexamethylenediamine*	3.39	3.93	4.12	0.409	0.522	0	0	0	2	4.055	5.676	N
201	3030-47-5	N-Methyl-N,N-bis(2-dimethylaminoethyl)amine	3.52	3.12	3.47	0.405	0.353	0	0	0	0	0.765	5.479	HPV
214	629-40-3	Octanedinitrile	3.05	2.83	3.26	0.452	0.338	0	0	0	2	1.301	6.491	N
216	107-13-1	2-propenenitrile	3.72	3.63	3.77	0.624	0.047	1	0	0	1	0.009	5.438	HPV
217	126-98-7	Methacrylonitrile	3.43	3.37	2.78	0.509	0.087	0	0	0	2	0.300	5.281	N
219	406-86-0	4,4,4-trifluorocrotonitrile	5.53	4.14	5.42	0.619	0.087	2	0	0	1	1.321	5.329	N
220	1855-63-6	1-cyclohexene-1-carbonitrile	3.42	3.91	3.44	0.440	0.204	1	0	0	2	1.679	5.049	N
221	1118-61-2	3-amino-2-Butenenitrile*	4.40	3.83	4.61	0.497	0.009	1	0	0	2	0.111	4.720	N
222	764-42-1	2-butenedinitrile, (e)-*	5.31	4.21	3.99	0.612	0.008	2	0	0	2	0.185	5.061	N
223	75-91-2	Tert-Butylhydroperoxide	4.91	3.10	3.47	0.404	0.219	0	0	0	0	0.535	5.129	HPV
224	3006-82-4	Tert-Butyl 2-ethylperoxyhexanoate*	5.86	4.90	4.36	0.372	0.610	0	0	0	0	10.429	4.884	HPV
228	96-29-7	2-Butanone oxime	3.74	3.28	3.68	0.476	0.189	0	0	0	1	0.344	5.322	HPV
229	100-64-1	Cyclohexanone oxime	4.06	3.56	3.97	0.434	0.301	0	0	0	2	0.963	5.316	HPV
234	57-14-7	N,N-dimethylhydrazine*	4.25	2.44	3.05	0.424	0.085	0	0	0	0	0.058	5.686	N
235	657-24-9	Metformin	2.61	3.44	2.99	0.431	0.020	0	2	0	8	0.108	4.669	N
236	110-91-8	Morpholine*	3.18	2.37	2.42	0.405	0.074	0	0	0	0	0.224	5.678	HPV
237	2403-88-5	2,2,6,6-tetramethylpiperidin-4-ol	3.12	3.03	3.10	0.343	0.310	0	0	0	1	1.738	5.523	HPV
239	110-85-0	Piperazine	2.82	2.31	2.42	0.393	0.141	0	0	0	0	0.224	5.876	HPV
240	108-80-5	Isocyanuric acid	2.13	3.09	2.78	0.471	-0.063	0	0	0	3	1.533	5.454	HPV
243	470-82-6	2-oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	2.79	3.09	2.87	0.346	0.378	0	0	0	0	6.262	6.016	N
246	15045-43-9	2,2,5,5-tetramethyltetrahydrofuran	2.34	2.84	2.39	0.360	0.385	0	0	0	0	3.541	6.075	N
247	62571-86-2	Captopril	3.11	4.64	3.88	0.384	0.127	0	0	1	2	0.410	4.412	N
248	674-82-8	But-3-en-3-olide	4.09	3.20	4.01	0.560	0.056	0	0	0	0	0.008	5.226	HPV
250	106-91-2	Glycidyl methacrylate	3.65	3.17	3.61	0.458	0.146	0	0	0	0	0.153	5.052	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
251	75-08-1	Ethanethiol	4.32	3.65	4.01	0.546	0.086	0	0	0	0	0.713	4.751	HPV
252	110-66-7	Pentane-1-thiol	5.87	5.30	5.79	0.509	0.305	0	0	1	0	4.957	4.750	N
253	111-88-6	1-mercaptooctane	7.02	6.97	7.17	0.496	0.548	0	0	1	0	17.664	4.749	HPV
254	143-10-2	1-decanethiol	7.35	7.79	7.17	0.490	0.701	0	0	1	0	23.266	4.749	N
256	624-92-0	Dimethyl disulphide	3.51	4.00	3.50	0.549	-0.019	0	0	0	0	0.713	4.032	HPV
257	110-81-6	Diethyl disulfide	4.44	4.22	4.36	0.464	0.119	0	0	0	0	3.301	4.045	N
259	3268-49-3	3-(methylthio)propionaldehyde	5.02	3.62	4.94	0.450	0.100	1	0	0	0	0.429	4.652	HPV
260	111-17-1	3,3'-thiodipropionic acid*	3.55	3.45	3.61	0.461	0.134	0	0	0	0	0.189	4.673	N
263	79-19-6	Thiosemicarbazide	3.68	3.59	3.65	0.513	-0.062	0	0	0	0	2.125	4.475	N
266	4189-44-0	Thiourea dioxide	3.34	4.24	3.38	0.549	-0.004	0	0	0	0	6.099	4.454	HPV
274	1763-23-1	Perfluorooctane sulfonic acid	4.16	4.90	4.34	0.412	0.066	0	0	1	0	15.048	5.532	N
285	115-96-8	Tris(2-chloroethyl) phosphate	2.80	3.09	2.86	0.415	0.077	0	0	0	0	5.559	5.483	HPV
294	71-43-2	Benzene	3.12	3.43	2.64	0.455	0.020	0	0	0	0	5.085	5.010	HPV
295	108-88-3	Toluene	3.53	3.80	3.82	0.450	0.071	0	0	0	0	6.802	4.846	HPV
296	100-41-4	Ethylbenzene	4.36	4.11	4.59	0.465	0.121	0	0	0	0	8.653	4.871	HPV
298	95-47-6	o-xylene	4.37	4.02	4.36	0.415	0.129	0	0	0	0	8.653	4.786	HPV
299	108-38-3	m-xylene*	4.08	4.04	4.68	0.434	0.114	0	0	0	0	8.653	4.816	HPV
300	106-42-3	p-xylene	3.83	4.16	3.87	0.439	0.124	0	0	0	0	8.653	4.695	HPV
301	103-65-1	n-Propylbenzene	4.82	4.48	4.59	0.465	0.206	0	0	0	0	10.620	4.864	HPV
302	98-82-8	Isopropylbenzene	4.66	4.32	4.66	0.418	0.215	0	0	0	0	10.620	4.883	HPV
303	104-51-8	Butylbenzene	4.92	4.98	4.91	0.471	0.257	0	0	0	0	12.690	4.630	N
304	99-87-6	p-Cymene	4.36	4.76	4.39	0.419	0.280	0	0	0	0	12.690	4.734	HPV
305	98-51-1	4-tert-butyltoluene	4.53	4.88	4.57	0.406	0.259	0	0	0	0	14.851	4.742	HPV
306	25321-09-9	Diisopropylbenzene	4.68	5.04	4.70	0.359	0.350	0	0	0	0	17.096	4.837	HPV
308	827-52-1	Cyclohexylbenzene	5.37	5.15	5.34	0.410	0.483	0	0	0	0	14.008	4.866	N
309	108-90-7	Chlorobenzene	3.58	4.20	3.67	0.535	0.002	0	0	0	0	8.272	4.737	HPV
310	95-49-8	2-chlorotoluene	4.21	4.39	4.26	0.484	0.066	0	0	0	0	10.302	4.674	HPV
311	108-41-8	3-chlorotoluene	4.55	4.36	4.50	0.504	0.033	0	0	0	0	10.302	4.717	N
312	106-43-4	4-chlorotoluene	4.45	4.49	4.50	0.514	0.044	0	0	0	0	10.302	4.622	HPV
314	95-50-1	1,2-dichlorobenzene	4.01	4.72	4.03	0.551	-0.020	0	0	0	0	12.094	4.562	HPV
315	541-73-1	1,3-dichlorobenzene	4.58	4.49	4.38	0.518	-0.049	0	0	0	0	12.094	4.632	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
316	106-46-7	1,4-dichlorobenzene	4.19	4.61	4.38	0.508	-0.019	0	0	0	0	12.094	4.514	HPV
317	95-73-8	2,4-dichlorotoluene	4.80	4.86	4.78	0.501	0.011	0	0	0	0	14.401	4.520	HPV
318	95-75-0	3,4-dichlorotoluene	4.93	5.00	4.87	0.538	0.008	0	0	0	0	14.401	4.502	N
319	19398-61-9	2,5-dichlorotoluene	4.98	4.80	5.14	0.465	0.035	0	0	0	0	14.401	4.494	N
320	118-69-4	2,6-dichlorotoluene	4.78	4.74	4.78	0.477	0.036	0	0	0	0	14.401	4.627	N
321	87-61-6	1,2,3-trichlorobenzene	5.05	5.09	4.81	0.553	-0.076	0	0	0	0	16.507	4.502	N
323	120-82-1	1,2,4-trichlorobenzene	4.57	5.11	4.81	0.535	-0.071	0	0	0	0	16.507	4.405	HPV
327	108-86-1	Bromobenzene*	4.12	4.44	4.19	0.588	-0.011	0	0	0	0	9.273	4.747	N
328	348-61-8	4-Bromo-1,2-difluorobenzene	4.14	4.96	4.19	0.555	-0.048	0	0	0	0	15.258	4.598	N
330	98-87-3	Alpha,alpha-dichlorotoluene	3.78	4.68	3.85	0.533	0.028	0	0	0	0	12.439	4.691	HPV
331	611-19-8	2-chlorobenzyl chloride	5.13	4.67	4.77	0.509	0.067	0	0	0	0	12.439	4.696	HPV
332	98-08-8	Benzotrifluoride ((trifluoromethyl)benzene)	4.43	4.44	4.77	0.520	0.051	0	0	0	0	11.358	4.862	HPV
333	402-31-3	Metaxylene hexafluoride*	4.44	5.01	5.38	0.448	0.071	0	0	0	0	19.270	4.859	N
334	98-83-9	2-phenylpropene	4.39	4.29	4.39	0.442	0.182	0	0	0	0	10.039	4.863	HPV
335	1321-74-0	Divinylbenzene	4.86	4.94	4.86	0.451	0.062	2	0	0	0	11.465	4.557	HPV
336	100-51-6	Benzyl alcohol	2.15	3.30	2.64	0.461	0.020	0	0	0	0	2.532	4.875	HPV
339	2100-42-7	2-chlorohydroquinonedimethylether	4.24	4.09	4.12	0.495	-0.059	0	0	0	0	4.688	4.079	HPV
342	93-15-2	4-allyl-1,2-dimethoxybenzene	3.91	4.37	3.99	0.446	0.066	1	0	0	0	5.768	4.254	N
343	122-57-6	Benzalacetone	5.42	4.67	5.58	0.490	0.102	2	0	0	0	5.612	4.447	N
345	100-52-7	Benzaldehyde	3.52	3.84	3.58	0.490	0.011	1	0	0	0	3.129	4.666	HPV
348	487-68-3	2,4,6-trimethylbenzaldehyde	4.27	4.33	4.26	0.422	0.123	1	0	0	0	7.414	4.555	N
352	123-11-5	p-methoxybenzaldehyde	3.35	3.95	3.39	0.504	-0.026	1	0	0	0	2.219	4.382	HPV
353	90-02-8	Salicylaldehyde/2-hydroxybenzaldehyde	4.41	4.02	4.19	0.479	-0.030	1	0	0	0	2.779	4.237	N
360	98-86-2	Acetophenone	5.08	3.78	4.32	0.467	0.096	0	0	0	0	4.420	4.712	HPV
363	84-66-2	Diethyl phthalate	3.69	3.71	3.71	0.418	0.044	0	0	0	0	6.641	4.741	HPV
364	84-69-5	Diisobutyl phthalate	5.19	4.52	5.14	0.360	0.252	0	0	0	0	13.090	4.758	HPV
365	84-74-2	Dibutyl phthalate*	5.01	4.65	4.70	0.377	0.287	0	0	0	0	13.090	4.751	HPV
366	131-17-9	Diallyl phthalate	4.22	4.56	4.29	0.406	0.137	2	0	0	0	8.703	4.684	HPV
368	99-04-7	m-toluic acid	3.90	3.74	3.84	0.474	0.001	0	0	0	0	4.135	4.525	HPV
369	99-94-5	4-methylbenzoic acid	3.26	3.67	3.26	0.485	-0.009	0	0	0	0	4.135	4.634	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
377	50-78-2	Acetylsalicylic acid*	3.23	3.61	2.64	0.452	0.022	0	0	0	0	2.894	4.483	HPV
379	69-72-7	Salicylic acid	3.33	3.63	3.38	0.483	-0.079	0	0	0	0	2.700	4.321	HPV
381	99-96-7	4-hydroxybenzoic acid	3.10	3.40	3.12	0.501	-0.091	0	0	0	0	1.282	4.492	HPV
382	19715-19-6	3,5-di-tert-butylsalicylic acid	4.57	5.17	4.60	0.350	0.232	0	0	0	0	15.692	4.151	N
394	2840-28-0	3-amino-4-chlorobenzoic acid	4.19	4.20	4.14	0.515	-0.148	0	0	0	2	3.006	3.983	N
397	NA	Gemfibrozil	4.22	4.50	4.32	0.357	0.295	0	0	0	0	10.127	4.496	N
398	94-74-6	2-methyl-4-chlorophenoxyacetic acid	2.62	3.66	2.73	0.437	-0.029	0	0	0	0	4.353	4.409	HPV
400	882-09-7	Clofibric acid	3.27	3.90	3.81	0.445	0.047	0	0	0	0	5.654	4.485	N
403	140-10-3	Trans-cinnamic acid	5.74	4.45	5.58	0.501	0.035	2	0	0	0	3.987	4.400	N
404	108-95-2	Phenol	2.73	3.46	2.75	0.477	-0.018	0	0	0	0	2.268	4.613	HPV
407	106-44-5	4-cresol	3.30	3.76	3.34	0.467	0.035	0	0	0	0	3.456	4.469	HPV
408	108-39-4	3-methylphenol*	2.87	3.59	2.64	0.462	0.020	0	0	0	0	3.456	4.627	HPV
410	95-48-7	2-methylphenol	2.93	3.64	3.03	0.441	0.060	0	0	0	0	3.456	4.571	HPV
411	90-00-6	2-ethylphenol	3.59	3.89	3.81	0.451	0.106	0	0	0	0	4.808	4.567	HPV
412	123-07-9	4-ethylphenol	3.75	3.99	3.77	0.479	0.079	0	0	0	0	4.808	4.495	HPV
413	620-17-7	3-ethylphenol*	3.48	3.84	3.29	0.470	0.066	0	0	0	0	4.808	4.620	HPV
414	526-75-0	2,3-dimethylphenol*	3.40	3.77	3.71	0.423	0.105	0	0	0	0	4.808	4.600	HPV
415	576-26-1	2,6-dimethylphenol	3.44	3.82	3.52	0.423	0.110	0	0	0	0	4.808	4.542	HPV
416	95-65-8	3,4-dimethylphenol	3.58	3.82	3.52	0.424	0.081	0	0	0	0	4.808	4.482	HPV
417	95-87-4	2,5-dimethylphenol*	3.60	3.93	3.82	0.445	0.113	0	0	0	0	4.808	4.512	HPV
418	105-67-9	2,4-dimethylphenol	3.96	3.97	3.81	0.444	0.108	0	0	0	0	4.808	4.434	HPV
419	108-68-9	3,5-dimethylphenol	3.65	3.64	3.65	0.428	0.065	0	0	0	0	4.808	4.692	HPV
422	527-60-6	2,4,6-trimethylphenol	4.15	4.12	4.21	0.415	0.163	0	0	0	0	6.299	4.437	HPV
423	697-82-5	2,3,5-trimethylphenol*	4.00	3.91	4.36	0.414	0.130	0	0	0	0	6.299	4.628	N
424	2416-94-6	2,3,6-trimethylphenol	3.98	4.00	4.00	0.410	0.144	0	0	0	0	6.299	4.523	HPV
425	88-18-6	2-tert-butyl phenol	5.06	4.06	5.00	0.396	0.130	0	0	0	0	7.914	4.563	HPV
428	89-72-5	O-sec-butylphenol	4.34	4.23	4.22	0.395	0.218	0	0	0	0	7.914	4.555	HPV
429	89-83-8	Thymol	4.03	4.35	4.22	0.418	0.251	0	0	0	0	7.914	4.578	HPV
430	99-71-8	p-sec-butylphenol	4.30	4.31	4.22	0.409	0.214	0	0	0	0	7.914	4.499	HPV
431	14938-35-3	4-pentylphenol*	4.87	4.86	4.91	0.480	0.277	0	0	0	0	9.639	4.490	N
432	88-60-8	6-tert-butyl-m-cresol	4.94	4.35	4.89	0.403	0.194	0	0	0	0	9.639	4.588	HPV
433	2219-82-1	6-tert-butyl-o-cresol	4.42	4.35	4.69	0.386	0.194	0	0	0	0	9.639	4.519	HPV
434	2409-55-4	2-tert-butyl-p-cresol	4.96	4.42	4.69	0.396	0.182	0	0	0	0	9.639	4.433	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
435	1879-09-0	2-(1,1-dimethylethyl)-4,6-dimethylphenol	4.40	4.68	4.32	0.374	0.255	0	0	0	0	11.463	4.412	N
436	96-76-4	2,4-di-tert-butylphenol*	5.31	4.95	4.70	0.358	0.254	0	0	0	0	15.375	4.481	HPV
437	1806-26-4	4-n-octylphenol	6.17	6.69	6.12	0.483	0.529	0	0	0	0	23.359	4.490	N
438	5510-99-6	2,6-di-sec-butylphenol	5.27	5.34	5.20	0.349	0.494	0	0	0	0	15.375	4.533	N
441	120-95-6	2,4-di-tert-pentylphenol	5.14	5.63	5.20	0.337	0.428	0	0	0	0	19.595	4.476	HPV
443	95-57-8	2-chlorophenol	3.39	3.92	3.63	0.519	-0.022	0	0	0	0	4.525	4.484	HPV
444	108-43-0	3-chlorophenol	4.05	3.78	3.69	0.500	-0.032	0	0	0	0	4.525	4.550	N
445	106-48-9	4-chlorophenol	3.86	4.01	3.63	0.537	-0.039	0	0	0	0	4.525	4.412	HPV
446	59-50-7	4-chloro-3-methylphenol	3.98	4.07	4.12	0.488	0.011	0	0	0	0	6.055	4.427	HPV
447	576-24-9	2,3-dichlorophenol	4.17	4.22	4.65	0.528	-0.048	0	0	0	0	7.446	4.441	N
448	120-83-2	2,4-dichlorophenol	4.42	4.18	4.36	0.501	-0.071	0	0	0	0	7.446	4.308	HPV
449	583-78-8	2,5-dichlorophenol	4.65	4.15	4.36	0.499	-0.049	0	0	0	0	7.446	4.395	HPV
450	87-65-0	2,6-dichlorophenol	4.01	4.19	4.36	0.513	-0.064	0	0	0	0	7.446	4.367	N
451	95-77-2	3,4-dichlorophenol	4.87	4.35	4.65	0.548	-0.058	0	0	0	0	7.446	4.330	N
452	591-35-5	3,5-dichlorophenol	4.89	4.08	4.65	0.531	-0.091	0	0	0	0	7.446	4.531	N
453	15950-66-0	2,3,4-trichlorophenol	4.68	4.58	4.59	0.521	-0.106	0	0	0	0	10.982	4.242	N
454	933-78-8	2,3,5-trichlorophenol*	4.94	4.48	4.70	0.518	-0.109	0	0	0	0	10.982	4.361	N
455	933-75-5	2,3,6-trichlorophenol	4.39	4.53	4.59	0.506	-0.094	0	0	0	0	10.982	4.273	N
456	95-95-4	2,4,5-trichlorophenol	4.77	4.62	4.59	0.527	-0.107	0	0	0	0	10.982	4.216	N
458	88-06-2	2,4,6-trichlorophenol	4.54	4.48	4.59	0.506	-0.142	0	0	0	0	10.982	4.212	HPV
459	58-90-2	2,3,4,6-tetrachlorophenol	5.04	4.72	5.00	0.517	-0.181	0	0	0	0	13.085	4.123	N
462	NA	Pentachlorophenol	5.49	4.86	4.51	0.522	-0.249	0	0	0	0	15.278	4.076	N
463	106-41-2	4-bromophenol	4.27	4.18	4.25	0.588	-0.061	0	0	0	0	5.272	4.459	N
464	615-58-7	2,4-dibromophenol	5.36	4.37	5.31	0.523	-0.086	0	0	0	0	9.277	4.357	N
465	118-79-6	2,4,6-tribromophenol	5.24	4.83	5.17	0.513	-0.113	0	0	0	0	14.171	4.266	HPV
468	1745-81-9	2-allylphenol	4.40	4.32	4.43	0.467	0.116	1	0	0	0	5.854	4.548	N
469	4286-23-1	4-(1-methylethenyl)phenol	4.40	3.97	3.81	0.456	0.058	0	0	0	0	5.854	4.499	N
470	90-05-1	2-methoxyphenol	2.66	3.47	2.72	0.469	-0.058	0	0	0	0	1.553	4.376	HPV
473	25013-16-5	Butylated hydroxyanisole	4.54	4.20	4.83	0.417	0.106	0	0	0	0	6.210	4.196	N
479	99-76-3	Methyl p-hydroxybenzoate	3.43	3.48	3.38	0.486	-0.061	0	0	0	0	2.148	4.505	N
480	88-75-5	2-nitrophenol	4.37	3.97	3.70	0.498	-0.081	0	0	0	2	1.816	4.215	HPV
483	89-64-5	4-chloro-2-nitrophenol	4.45	4.16	4.94	0.465	-0.111	0	0	0	2	3.876	4.009	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
487	88-30-2	3-Trifluoromethyl-4-nitrophenol	4.93	4.53	4.49	0.485	-0.013	0	0	0	3	6.065	4.342	N
507	121-79-9	Propyl gallate	3.16	3.73	3.16	0.485	-0.015	0	0	0	0	1.054	4.157	N
508	62-53-3	Aniline	3.09	3.66	3.27	0.459	-0.102	0	0	0	2	2.268	4.450	HPV
509	95-53-4	2-methylaniline	2.97	4.01	3.02	0.428	-0.017	0	0	0	3	3.456	4.417	HPV
510	108-44-1	3-methylaniline	3.60	3.78	3.57	0.447	-0.066	0	0	0	2	3.456	4.489	HPV
511	106-49-0	4-methylaniline	3.53	3.94	3.57	0.453	-0.049	0	0	0	2	3.456	4.339	HPV
512	578-54-1	2-ethylaniline	3.39	4.26	3.43	0.439	0.024	0	0	0	3	4.808	4.413	N
513	587-02-0	3-ethylaniline	3.93	4.05	3.91	0.458	-0.016	0	0	0	2	4.808	4.477	N
514	589-16-2	4-ethylaniline	4.14	4.20	3.83	0.466	0.001	0	0	0	2	4.808	4.362	N
515	87-59-2	2,3-dimethylaniline	3.53	4.20	3.57	0.412	0.062	0	0	0	3	4.808	4.461	N
516	87-62-7	2,6-dimethylaniline	3.05	4.38	3.15	0.412	0.056	0	0	0	4	4.808	4.412	HPV
517	95-64-7	3,4-dimethylaniline	4.19	4.04	3.90	0.426	-0.005	0	0	0	2	4.808	4.372	N
518	95-68-1	2,4-dimethylaniline	3.49	4.32	3.43	0.433	0.031	0	0	0	3	4.808	4.327	HPV
519	95-78-3	2,5-dimethylaniline	3.42	4.23	3.43	0.432	0.026	0	0	0	3	4.808	4.422	N
520	108-69-0	3,5-dimethylaniline	3.63	3.82	3.90	0.417	-0.034	0	0	0	2	4.808	4.545	N
521	579-66-8	2,6-diethylaniline	3.56	4.87	4.84	0.409	0.153	0	0	0	4	7.914	4.394	HPV
522	99-88-7	4-isopropylaniline	3.88	4.34	3.88	0.424	0.080	0	0	0	2	6.299	4.377	HPV
523	88-05-1	2,4,6-trimethylaniline	3.76	4.62	4.50	0.405	0.084	0	0	0	4	6.299	4.332	N
525	95-51-2	2-chloroaniline*	3.59	4.15	4.48	0.501	-0.093	0	0	0	2	4.525	4.318	HPV
526	108-42-9	3-chloroaniline	3.76	4.08	3.81	0.516	-0.121	0	0	0	2	4.525	4.400	N
527	106-47-8	4-chloroaniline*	4.55	4.20	4.58	0.527	-0.124	0	0	0	2	4.525	4.286	N
528	95-81-8	2-chloro-5-methylaniline	4.46	4.30	4.49	0.493	-0.063	0	0	0	2	6.055	4.349	N
529	95-76-1	3,4-dichloroaniline	4.61	4.58	4.58	0.552	-0.146	0	0	0	2	7.446	4.221	HPV
530	95-82-9	2,5-dichloroaniline	4.11	4.39	3.98	0.481	-0.106	0	0	0	2	7.446	4.252	N
531	554-00-7	2,4-dichloroaniline*	4.61	4.49	4.81	0.511	-0.142	0	0	0	2	7.446	4.169	N
532	608-27-5	2,3-dichloroaniline	4.38	4.43	4.48	0.510	-0.130	0	0	0	2	7.446	4.267	N
533	608-31-1	2,6-dichloroaniline	3.84	4.50	3.98	0.494	-0.091	0	0	0	2	7.446	4.205	N
534	626-43-7	3,5-dichloroaniline	4.57	4.33	4.48	0.521	-0.164	0	0	0	2	7.446	4.363	N
535	634-67-3	2,3,4-trichloroaniline	4.74	4.89	4.94	0.538	-0.183	0	0	0	2	10.982	4.111	N
536	634-93-5	2,4,6-trichloroaniline	4.67	4.79	4.74	0.487	-0.164	0	0	0	2	10.982	4.067	N
537	636-30-6	2,4,5-trichloroaniline	4.80	4.87	4.81	0.517	-0.160	0	0	0	2	10.982	4.097	N
538	634-91-3	3,4,5-trichloroaniline	5.14	4.87	4.94	0.552	-0.196	0	0	0	2	10.982	4.168	N
544	104-94-9	p-anisidine*	3.98	3.86	3.52	0.483	-0.135	0	0	0	2	1.553	4.123	HPV
545	88-74-4	2-nitroaniline	3.50	4.25	3.56	0.482	-0.121	0	0	0	4	1.816	4.073	HPV
546	99-09-2	3-nitroaniline*	3.51	4.33	3.24	0.496	-0.154	0	0	0	4	1.816	3.951	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
547	100-01-6	4-nitroaniline	3.51	4.16	3.24	0.509	-0.175	0	0	0	4	1.816	4.172	HPV
548	99-52-5	2-methyl-4-nitroaniline*	4.20	4.46	4.63	0.458	-0.100	0	0	0	5	2.893	4.080	N
552	89-63-4	4-chloro-2-nitroaniline	4.31	4.48	4.51	0.478	-0.157	0	0	0	4	3.876	3.923	HPV
554	96-96-8	2-nitro-p-anisidine	4.15	4.42	4.32	0.485	-0.159	0	0	0	4	1.395	3.712	N
558	103-69-5	N-ethylaniline	3.56	4.27	3.83	0.458	0.034	0	0	0	2	4.808	4.312	HPV
559	121-69-7	N,n-dimethylaniline	3.74	4.15	3.72	0.430	-0.015	0	0	0	2	4.808	4.228	HPV
560	91-66-7	N,n-diethylaniline	4.73	4.64	4.70	0.402	0.142	0	0	0	2	7.914	4.249	HPV
561	106-50-3	p-phenylenediamine	5.78	3.89	5.60	0.460	-0.211	0	0	0	4	0.798	4.083	HPV
562	108-45-2	m-phenylenediamine	3.56	3.61	4.35	0.455	-0.206	0	0	0	4	0.798	4.444	HPV
563	95-54-5	o-phenylenediamine	5.12	3.77	4.35	0.441	-0.170	0	0	0	4	0.798	4.253	HPV
564	95-70-5	2,5-diaminotoluene	4.79	4.16	4.31	0.438	-0.140	0	0	0	5	1.553	4.103	N
565	95-80-7	2,4-diaminotoluene	3.83	4.02	4.31	0.438	-0.129	0	0	0	5	1.553	4.321	HPV
570	101-96-2	1,4-benzenediamine, N,N'-bis(1-methylpropyl)-	5.37	5.79	5.43	0.375	0.142	0	0	0	6	10.945	3.795	HPV
571	85068-29-7	3,5-Bis(trifluoromethyl)benzylamine	4.83	5.11	4.84	0.429	0.105	0	0	0	2	13.350	4.363	N
572	1477-55-0	m-phenylenebis (methylamine)	3.69	3.77	3.66	0.409	0.055	0	0	0	4	1.090	4.710	HPV
575	29122-68-7	Atenolol	2.63	3.59	2.69	0.432	0.220	0	1	0	2	0.855	4.479	HPV
578	591-27-5	3-aminophenol	2.83	3.51	2.94	0.471	-0.125	0	0	0	2	0.798	4.457	HPV
579	119-34-6	4-amino-2-nitrophenol	4.49	4.35	4.32	0.492	-0.186	0	0	0	4	0.685	3.684	N
580	98-95-3	Nitrobenzene*	3.64	3.97	3.81	0.513	-0.042	0	0	0	2	3.562	4.613	HPV
581	88-72-2	2-nitrotoluene	3.79	4.37	3.83	0.468	0.039	0	0	0	3	5.020	4.466	HPV
583	99-99-0	4-methylnitrobenzene*	4.14	4.17	3.81	0.498	-0.013	0	0	0	2	5.020	4.538	HPV
585	88-73-3	2-chloronitrobenzene	3.92	4.45	3.98	0.521	-0.040	0	0	0	2	6.293	4.367	HPV
586	121-73-3	3-chloronitrobenzene*	4.09	4.29	4.48	0.486	-0.083	0	0	0	2	6.293	4.319	HPV
587	100-00-5	4-chloronitrobenzene	4.10	4.28	4.49	0.489	-0.061	0	0	0	2	6.293	4.397	HPV
588	13290-74-9	4-chloro-3-methylnitrobenzene*	4.72	4.54	4.96	0.458	0.002	0	0	0	2	8.078	4.306	N
591	99-54-7	3,4-dichloronitrobenzene	4.89	4.78	4.86	0.523	-0.114	0	0	0	2	9.673	4.198	HPV
593	89-69-0	1,2,4-trichloro-5-nitrobenzene*	5.18	5.12	4.73	0.498	-0.132	0	0	0	2	13.655	4.098	HPV
597	350-30-1	2-chloro-1-fluoro-4-nitrobenzene*	5.47	4.57	4.81	0.504	-0.097	0	0	0	2	8.676	4.303	N
599	91-23-6	2-nitroanisole	3.43	3.95	3.27	0.457	-0.045	0	0	0	2	2.893	4.281	HPV
600	555-03-3	3-nitroanisole	3.93	4.19	3.92	0.497	-0.063	0	0	0	2	2.893	4.107	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
608	606-20-2	2,6-dinitrotoluene	4.08	4.59	4.16	0.430	-0.001	0	0	0	6	5.014	4.489	N
609	97-00-7	1-chloro-2,4-dinitrobenzene	6.05	4.54	6.37	0.498	-0.118	0	0	0	4	6.286	4.348	HPV
614	534-52-1	4,6-dinitro-o-cresol	4.55	4.23	4.49	0.445	-0.073	0	0	0	4	3.055	4.209	N
616	88-85-7	2-(1-methylpropyl)-4,6-dinitro-phenol (Dinoseb)	5.23	4.80	4.50	0.402	0.090	0	0	0	4	7.301	4.225	HPV
621	40487-42-1	Pendimethalin	7.30	6.54	6.62	0.368	0.263	0	0	0	9	12.577	3.885	HPV
622	1582-09-8	Trifluralin*	6.18	6.90	6.62	0.373	0.321	0	0	0	8	17.698	4.037	HPV
623	55283-68-6	Ethalfluraline	5.77	6.76	6.62	0.372	0.289	0	0	0	8	16.946	4.046	N
624	29091-05-2	Dinitramine	7.02	5.66	6.92	0.389	0.049	0	0	0	9	8.592	4.075	N
628	873-32-5	o-chlorobenzonitrile	3.86	4.23	3.81	0.528	-0.076	0	0	0	2	5.618	4.513	N
630	91-15-6	Phthalonitrile	2.96	3.83	3.24	0.504	-0.166	0	0	0	4	1.970	4.628	HPV
633	23950-58-5	Propyzamide	4.88	5.22	4.96	0.476	0.007	0	0	0	3	14.486	4.522	N
635	51218-45-2	Metolachlor	6.60	4.87	6.42	0.358	0.304	0	0	0	4	9.155	4.700	HPV
641	57837-19-1	Methyl-(2-methoxyacetyl)-N-(2,6-xylyl)-DL-alaninate	4.65	4.07	4.63	0.374	0.143	0	0	0	4	3.636	4.709	N
642	103-90-2	Paracetamol	3.05	3.99	3.70	0.490	-0.024	0	0	0	2	1.123	4.217	HPV
643	3766-81-2	Fenobucarb*	3.80	4.20	4.21	0.403	0.271	0	0	0	1	5.716	4.675	N
644	114-26-1	Propoxur	4.32	3.75	4.27	0.412	0.061	0	0	0	1	2.465	4.360	HPV
648	34123-59-6	Isoproturon/1,1-dimethyl-3-(8-isopropylphenyl)-urea	6.09	4.93	4.84	0.423	0.195	0	0	0	5	5.716	4.405	HPV
651	330-54-1	1-(3,4 dichlorophenyl)-3,3 dimethyl urea	6.70	5.02	6.37	0.504	-0.031	0	0	0	5	7.000	4.255	HPV
654	5329 12 4	2,4,6-trichlorophenyl hydrazine	6.40	5.16	5.99	0.477	-0.122	0	0	0	4	10.531	3.968	HPV
658	108-98-5	Benzenethiol	5.84	4.28	5.76	0.512	0.033	0	0	0	0	6.365	4.349	N
659	28249-77-6	Thiobencarb	6.45	5.77	5.46	0.418	0.251	0	0	1	1	9.563	4.376	N
661	88-19-7	o-toluenesulfonamide	3.00	2.77	3.04	0.458	0.065	0	2	0	2	0.606	4.677	HPV
662	63-74-1	Sulphanilamide	3.87	2.99	3.83	0.497	-0.108	0	2	0	4	0.013	4.447	N
664	98-59-9	4-toluenesulfonyl chloride	3.52	4.65	3.68	0.527	-0.011	0	0	1	0	3.793	4.747	N
667	121-03-9	4-nitrotoluene-2-sulphonic acid	3.26	3.42	3.28	0.446	-0.095	0	0	0	2	1.399	4.612	HPV
670	15318-45-3	Thiamphenicol	4.60	4.54	3.88	0.429	0.197	0	0	1	1	0.634	4.759	N
672	122-14-5	Fenitrothion	5.75	4.90	5.72	0.434	-0.093	0	0	1	3	3.684	4.365	N
677	26087-47-8	Iprobenfos	4.49	5.06	5.46	0.382	0.154	0	0	1	0	7.844	4.477	N
686	10500-57-9	5,6,7,8-tetrahydroquinoline	3.32	4.34	3.44	0.418	0.311	0	0	0	3	3.422	4.756	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
687	100-43-6	4-vinylpyridine	4.36	3.84	4.31	0.517	0.071	1	0	0	1	1.229	4.904	N
688	100-69-6	2-vinylpyridine	3.23	4.39	3.27	0.515	0.140	1	0	0	2	1.229	4.536	HPV
693	504-24-5	4-aminopyridine	3.50	3.63	3.52	0.478	-0.067	0	0	0	3	0.021	4.577	N
694	504-29-0	2-aminopyridine	3.89	3.38	3.81	0.475	-0.019	0	1	0	3	0.445	4.485	N
700	122-34-9	Simazine	5.96	5.89	6.21	0.485	0.237	0	0	0	11	5.165	4.641	HPV
701	1912-24-9	Atrazine	6.08	6.04	6.21	0.468	0.266	0	0	0	11	6.707	4.643	HPV
704	5915-41-3	Terbuthylazine*	6.80	6.17	6.21	0.452	0.281	0	0	0	11	8.371	4.643	N
708	33693-04-8	Terbumeton	6.58	6.16	6.21	0.415	0.356	0	0	0	13	6.387	4.820	N
710	NA	2-methylthio-4-tert-butylamino-6-amino-s-triazine	6.67	5.78	7.38	0.423	0.158	0	2	1	11	5.165	4.395	N
711	1014-70-6	Simetryn	6.88	7.10	7.38	0.450	0.187	0	0	1	13	5.165	4.414	N
715	886-50-0	Terbutryn	7.86	7.52	7.38	0.420	0.305	0	0	1	13	8.371	4.428	N
719	28159-98-0	Irgarol 1051	8.04	7.59	7.38	0.400	0.279	0	0	1	13	10.142	4.403	N
722	51-21-8	5-fluorouracil	3.43	3.81	3.44	0.489	-0.032	1	0	0	2	0.055	4.598	N
724	21087-64-9	Metribuzin	6.74	6.53	6.76	0.430	0.233	0	0	1	9	2.556	4.054	N
726	110-02-1	Thiophene	2.88	4.03	2.94	0.534	0.165	0	0	0	0	1.375	4.472	N
729	443-48-1	Metronidazole	3.63	3.73	3.66	0.423	0.025	0	0	0	3	0.196	4.448	N
730	61-82-5	3-amino-1,2,4-triazole	3.11	2.69	3.14	0.518	-0.079	0	1	0	1	0.054	4.989	HPV
736	92-52-4	Biphenyl	5.30	4.66	5.14	0.459	-0.031	0	0	0	0	15.023	4.558	HPV
737	5707-44-8	4-ethyl-1,1'-biphenyl*	6.08	5.37	5.38	0.463	0.075	0	0	0	0	19.460	4.491	N
739	90-43-7	2-phenylphenol	4.53	4.06	4.69	0.453	-0.118	0	0	0	0	9.778	4.428	HPV
741	92-69-3	p-phenylphenol	4.85	4.25	4.69	0.470	-0.064	0	0	0	0	9.778	4.394	N
742	92-88-6	4,4'-dihydroxy-biphenyl	4.51	3.92	4.49	0.473	-0.097	0	0	0	0	6.321	4.320	HPV
743	119-93-7	o-tolidine	4.53	4.95	4.57	0.416	-0.157	0	0	0	6	9.300	4.110	N
744	91-94-1	3,3'-dichlorobenzidine	5.26	4.88	5.34	0.462	-0.328	0	0	0	4	12.856	4.045	HPV
745	58-14-0	Pyrimethamine	4.69	4.33	4.04	0.438	-0.035	0	3	0	10	9.293	4.388	N
746	91-76-9	2,4-Diamino-6-phenyl-s-triazine/benzoguanamine	3.42	3.68	4.04	0.476	-0.061	0	4	0	11	5.434	4.488	HPV
747	1698-60-8	Chloridazon	4.64	5.27	4.65	0.495	-0.049	1	0	0	7	3.809	4.129	HPV
751	51963-82-7	Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-*	4.56	4.14	4.64	0.364	-0.033	0	0	0	4	1.043	3.802	N
752	32809-16-8	Procymidone	5.70	5.23	5.62	0.436	0.111	0	0	0	5	10.500	4.472	N
753	18854-01-8	Isoxathion	5.56	5.18	5.55	0.459	-0.077	0	0	1	1	7.867	4.286	N
757	19666-30-9	Oxadiazon	6.79	6.53	6.62	0.401	0.167	0	0	0	8	16.237	4.074	N
758	147-94-4	Cytarabine	3.66	4.27	3.72	0.420	-0.009	2	1	0	7	1.382	4.623	N
760	95058-81-4	Gemcitabine	3.77	4.27	3.72	0.441	0.015	2	1	0	7	0.015	4.602	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
762	96-09-3	Styrene-7,8-oxide*	3.68	3.35	2.64	0.466	0.029	0	0	0	0	2.349	4.838	N
763	901-44-0	2,2-Bis[4-(2-hydroxyethoxy)phenyl]propane	4.27	4.16	4.21	0.408	0.170	0	0	0	0	6.889	4.445	N
764	599-64-4	4-(α,α-dimethylbenzyl)phenol	5.18	4.88	5.16	0.400	0.154	0	0	0	0	15.355	4.514	HPV
765	620-92-8	4,4'-dihydroxydiphenylmethane	4.10	4.23	3.82	0.442	0.093	0	0	0	0	7.762	4.433	N
766	80-05-7	Bisphenol A	4.68	4.41	4.69	0.402	0.111	0	0	0	0	10.928	4.463	HPV
767	79-94-7	Tetrabromobisphenol A	4.76	5.58	4.84	0.402	-0.204	0	0	0	0	27.612	4.274	HPV
768	101-77-9	4,4'-methylenedianiline	4.22	4.58	4.16	0.437	-0.074	0	0	0	4	7.762	4.313	HPV
771	101-84-8	Diphenyl ether	5.47	4.41	5.17	0.460	-0.031	0	0	0	0	11.526	4.448	HPV
776	101-80-4	4,4'-diaminodiphenyl ether	3.85	4.11	3.93	0.461	-0.318	0	0	0	4	4.904	4.056	HPV
782	103-50-4	Dibenzyl ether	4.68	4.18	4.68	0.433	0.059	0	0	0	0	11.522	4.854	N
785	119-61-9	Benzophenone	4.89	4.42	5.17	0.452	-0.007	0	0	0	0	12.867	4.628	HPV
786	NA	Fenofibrate	4.26	5.49	5.05	0.399	0.238	0	0	0	0	18.197	4.287	N
788	131-57-7	2-hydroxy-4-methoxybenzophenone	5.53	4.42	5.41	0.464	-0.014	0	0	0	0	10.119	4.319	N
792	122-39-4	Diphenylamine	5.60	5.14	5.99	0.454	-0.096	0	0	0	4	11.526	4.090	HPV
793	620-93-9	Di-p-tolylamine*	6.15	5.67	6.43	0.444	-0.029	0	0	0	4	15.448	4.019	N
800	101-20-2	3,4,4'-trichlorodiphenylurea*	6.82	5.79	6.43	0.518	-0.227	0	0	0	6	15.900	4.151	N
801	102-06-7	1,3-diphenylguanidine*	4.44	4.88	4.04	0.464	-0.206	0	1	0	8	9.538	4.144	HPV
802	97-39-2	N,N'-Bis(2-methylphenyl) guanidine	4.44	5.51	4.55	0.424	-0.104	0	1	0	10	13.019	4.239	N
803	122-66-7	Hydrazobenzene	5.22	5.49	5.20	0.445	-0.084	0	0	0	8	11.064	4.348	N
806	60-09-3	P-Aminoazobenzene	4.83	5.80	4.90	0.486	-0.169	0	0	0	10	7.175	3.799	N
807	80-09-1	Bis(4-hydroxyphenyl) Sulfone	3.59	3.64	4.32	0.454	0.005	0	0	0	0	4.249	4.587	HPV
810	30171-80-3	Dibromocresyl glycidyl ether	5.72	4.48	5.76	0.513	0.042	0	0	0	0	7.037	4.206	N
816	738-70-5	Trimethoprim	3.35	3.37	2.99	0.407	-0.015	0	3	0	9	1.591	4.384	N
818	68-35-9	Sulfadiazine	4.51	5.17	4.51	0.450	-0.162	0	0	1	8	0.237	4.443	N
819	57-68-1	Sulfamethazine	4.51	5.30	4.51	0.423	-0.092	0	0	1	8	1.123	4.447	N
823	122-11-2	Sulfadimethoxine	4.44	4.64	4.43	0.433	-0.237	0	0	0	11	2.086	4.450	N
829	64902-72-3	Chlorsulfuron*	6.32	5.43	5.35	0.450	-0.031	0	0	0	12	4.492	4.518	N
836	77732-09-3	Oxadixyl	4.50	4.37	4.16	0.366	0.023	0	0	0	8	1.687	4.515	N
842	723-46-6	Sulfamethoxazole	5.12	4.98	5.12	0.458	-0.159	0	0	1	6	0.942	4.439	N

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
855	90-12-0	1-methylnaphthalene	4.71	4.80	5.06	0.428	0.028	0	0	0	0	13.524	4.191	N
856	91-57-6	2-methylnaphthalene	4.87	4.78	5.00	0.447	-0.011	0	0	0	0	13.524	4.209	N
857	573-98-8	1,2-dimethylnaphthalene*	4.58	5.10	5.33	0.418	0.074	0	0	0	0	15.670	4.150	N
858	575-41-7	1,3-dimethylnaphthalene	5.40	5.03	5.06	0.418	0.047	0	0	0	0	15.670	4.171	N
859	582-16-1	2,7-dimethylnaphthalene	5.19	4.98	5.18	0.434	0.011	0	0	0	0	15.670	4.218	N
860	29253-36-9	Isopropylnaphthalene	5.83	5.37	5.05	0.402	0.149	0	0	0	0	17.896	4.190	N
861	525-66-6	Propranolol	5.13	4.45	4.83	0.409	0.123	0	0	0	1	6.420	4.093	N
863	135-19-3	B-naphthol*	4.84	4.08	4.69	0.461	-0.106	0	0	0	0	6.954	4.106	HPV
865	91-59-8	B-naphthylamine	5.46	4.32	5.44	0.457	-0.184	0	0	0	2	6.954	3.969	N
866	479-27-6	1,8-naphthylenediamine	5.52	4.51	5.19	0.418	-0.205	0	0	0	6	4.097	3.913	N
867	2243-62-1	1,5-naphthalenediamine	4.88	4.53	5.19	0.422	-0.227	0	0	0	6	4.097	3.848	HPV
868	22204-53-1	Naproxen	3.81	4.40	3.79	0.460	0.005	0	0	0	0	7.605	4.045	HPV
869	92-70-6	3-Hydroxy-2-naphthoic acid	3.47	4.22	3.54	0.472	-0.182	0	0	0	0	7.400	3.840	HPV
871	58-27-5	2-methyl-1,4-naphthoquinone	5.87	4.41	5.75	0.448	0.141	1	0	0	0	4.987	4.294	N
872	117-80-6	2,3-Dichloro-1,4-naphthoquinone	6.50	4.88	6.35	0.523	0.021	0	0	0	0	9.304	3.968	N
873	1785-65-5	2-acetoxy-1,4-naphthoquinone	4.70	4.30	4.69	0.474	0.106	1	0	0	0	3.500	4.289	N
874	83-72-7	2-hydroxy-1,4-naphthoquinone	3.97	4.07	4.19	0.467	-0.031	1	0	0	0	1.888	4.000	N
875	91-22-5	Quinoline*	3.29	4.24	3.91	0.460	-0.026	0	0	0	3	4.267	4.339	HPV
876	91-53-2	6-Ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	5.49	5.53	6.23	0.393	0.212	1	0	0	3	7.986	3.728	N
877	148-24-3	8-hydroxyquinoline	5.45	4.39	4.94	0.461	-0.053	0	0	0	3	3.857	4.032	N
883	22720-75-8	1-Benzo[b]thien-2-ylethan-1-one*	4.51	4.81	6.35	0.487	0.119	0	0	0	0	6.954	3.840	N
888	95-31-8	N-(tert-Butyl)-2-benzothiazolylsulfenamide	6.40	5.75	6.29	0.444	0.223	0	0	1	2	4.552	4.006	HPV
889	95-33-0	N-Cyclohexyl-2-benzothiazolylsulfenamide	6.25	6.31	6.17	0.436	0.345	0	0	1	4	5.264	4.021	HPV
890	149-30-4	2-mercaptobenzothiazole	5.52	5.41	5.49	0.536	-0.138	0	0	1	2	3.129	3.771	HPV
894	85-44-9	Phthalic anhydride	3.34	3.87	3.69	0.503	0.022	0	0	0	0	4.804	4.622	HPV
895	117-08-8	Tetrachlorophthalic anhydride	3.53	5.06	4.51	0.502	-0.144	0	0	0	0	15.374	3.996	HPV

Table B.1. Continued.

ID	CAS	Name	Exp pEC_{50}	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV[a] Status
900	83-32-9	Acenaphthene*	5.04	4.96	5.37	0.415	0.024	0	0	0	0	15.670	4.191	HPV
901	85-01-8	Phenanthrene	5.44	5.10	5.38	0.440	-0.134	0	0	0	0	18.762	4.118	N
902	86-73-7	Fluorene*	5.34	5.27	5.38	0.444	0.069	0	0	0	0	17.204	4.222	N
916	1484-13-5	9-vinylcarbazole	6.96	5.43	6.23	0.421	-0.053	1	0	0	4	10.986	3.905	N
917	298-46-4	Carbamazepine	3.50	4.34	3.49	0.407	-0.282	2	2	0	8	6.701	4.092	HPV
921	2222-33-5	Dibenzo[b,f]cyclohepten-1-one	6.17	5.23	6.11	0.431	-0.006	2	0	0	0	14.146	4.266	N
923	132-65-0	Dibenzothiophene	5.12	5.09	5.00	0.467	-0.071	0	0	0	0	15.023	3.940	N
924	1916-55-8	2-acetamidophenoxazin-3-one*	5.77	5.45	5.97	0.471	0.005	2	0	0	6	0.226	3.571	N
925	1916-59-2	2-aminophenoxazin-3-one	6.14	5.32	5.97	0.473	-0.096	2	0	0	6	0.159	3.490	N
926	NA	2-amino-7-methoxyphenoxazin-3-one*	6.66	5.37	5.97	0.484	-0.106	2	0	0	6	0.023	3.442	N
927	92-84-2	Phenothiazine	5.43	5.23	5.46	0.459	-0.046	0	0	0	4	8.174	3.693	HPV
928	14698-29-4	Oxolinic acid	4.21	4.37	4.26	0.428	-0.076	1	0	0	4	0.893	3.975	N
929	42835-25-6	Flumequine	4.72	5.01	4.64	0.407	0.055	1	0	0	5	4.941	4.082	N
933	2439-01-2	Chinomethionate	6.38	5.62	6.32	0.533	-0.114	0	0	1	6	0.199	3.959	N
935	90-30-2	1-(N-phenylamino)-naphthalene	6.81	6.05	6.43	0.445	-0.153	0	0	0	5	18.306	3.773	HPV
936	135-88-6	N-phenyl-2-naphthylamine	6.08	5.73	6.43	0.443	-0.227	0	0	0	4	18.306	3.801	N
937	88426-33-9	Buparvaquone	5.89	6.26	5.86	0.386	0.746	0	0	0	0	14.130	3.983	N
938	79617-96-2	Sertraline	6.39	6.76	6.27	0.405	0.240	0	0	0	3	26.513	4.308	N
941	70458-96-7	Norfloxacin	4.28	4.79	4.64	0.401	-0.026	1	0	0	6	2.046	3.968	N
942	85721-33-1	Ciprofloxacin	4.69	4.81	4.64	0.398	-0.067	1	0	0	6	2.791	3.925	N
943	93106-60-6	Enrofloxacin	4.10	5.30	4.64	0.394	0.088	1	0	0	6	4.560	3.870	N
944	98079-51-7	Lomefloxacin*	5.17	5.53	4.53	0.398	0.161	1	0	0	7	3.199	3.793	N
946	948-65-2	2-phenylindole	5.99	5.28	5.99	0.466	-0.125	0	0	0	4	12.835	4.055	N
956	115-86-6	Triphenyl phosphate	4.91	5.08	4.84	0.407	-0.095	0	0	0	0	22.938	4.627	HPV
958	27955-94-8	Phenol,4,4',4''-ethylidynetris-	4.90	4.54	4.89	0.379	0.081	0	0	0	0	13.653	4.464	N
979	60-54-8	Tetracycline	5.65	4.67	5.33	0.359	0.396	0	1	0	5	1.922	3.900	N
981	79-57-2	Oxytetracycline	5.01	4.44	5.33	0.348	0.187	0	1	0	5	4.452	3.950	N
986	82419-36-1	Ofloxacin	5.40	5.10	4.64	0.397	0.068	1	0	0	6	1.314	3.689	N
993	95233-18-4	Atovaquone*	6.19	6.24	6.12	0.427	0.430	0	0	0	0	17.873	3.876	N
1005	23696-28-8	Olaquindox	3.82	4.98	4.16	0.405	0.037	0	0	0	7	0.098	3.534	N

*Test set compound. [a]Production volume status according to OECD (2009). HPV: High production volume. N: Not HPV.

Table B.2. Algae external set chemicals, predicted pEC_{50}, model descriptors, production volume status.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
58-89-9	Lindane	4.23	4.68	0.419	-0.038	0	0	0	0	16.744	5.135	HPV
75-35-4	1,1-dichloroethylene	3.87	4.25	0.655	0.044	0	0	0	0	2.795	5.110	HPV
67-56-1	Methanol	1.95	1.82	0.502	0.016	0	0	0	0	0.663	6.585	HPV
67-63-0	2-propanol	1.88	1.82	0.430	0.152	0	0	0	0	0.120	6.620	HPV
20679-58-7	Acetic acid, bromo-, 2-butene-1,4-diyl ester	4.51	4.61	0.520	0.192	2	0	0	0	3.749	4.775	N
80-62-6	Methyl methacrylate	3.03	3.13	0.470	0.095	0	0	0	0	0.789	5.235	HPV
75-64-9	T-Butylamine	1.98	2.58	0.371	0.221	0	0	0	0	0.640	6.455	HPV
124-40-3	Dimethylamine	2.31	1.82	0.463	0.096	0	0	0	0	0.030	6.055	HPV
108-18-9	Diisopropylamine	2.87	3.05	0.394	0.321	0	0	0	0	2.519	5.895	HPV
111-92-2	Dibutylamine	3.98	3.23	0.460	0.496	0	0	0	2	5.170	5.910	HPV
68-12-2	Dimethylformamide	3.13	4.31	0.459	0.097	1	0	0	0	0.074	5.275	HPV
62-75-9	Dimethylnitrosamine	3.24	3.13	0.476	0.108	0	0	0	0	0.168	4.950	N
55-18-5	Diethylnitrosamine	3.73	3.14	0.444	0.237	0	0	0	2	0.316	4.885	N
99129-21-2	Clethodim	7.24	6.17	0.397	0.641	2	0	1	4	5.977	4.080	N
62-56-6	Thiourea	3.79	4.01	0.553	-0.039	0	0	0	0	2.056	4.445	HPV
2212-67-1	Molinate	5.38	3.88	0.439	0.305	0	0	1	3	2.494	4.600	HPV
77182-82-2	Glufosinate	3.67	2.86	0.426	0.094	0	0	0	1	8.680	5.410	N
1071-83-6	Glyphosate	3.55	3.47	0.508	0.018	0	0	0	0	3.840	4.930	HPV
126-72-7	Tris-(2,3-dibromopropyl) phoshate	5.42	4.89	0.355	0.220	0	0	0	0	22.613	4.695	N
115-29-7	Endosulfan	5.63	5.46	0.408	0.343	0	0	1	0	9.706	4.555	HPV
131-11-3	Dimethyl phthalate	3.30	2.64	0.425	-0.035	0	0	0	0	4.018	4.765	HPV
644-35-9	2-n-propylphenol	4.22	4.32	0.456	0.184	0	0	0	0	6.299	4.550	N
98-54-4	P-tert-butylphenol	4.17	3.65	0.411	0.148	0	0	0	0	7.914	4.520	HPV
104-40-5	4-n-Nonylphenol	7.09	4.91	0.482	0.610	0	0	0	0	25.901	4.490	N
609-19-8	3,4,5-trichlorophenol	4.66	3.38	0.563	-0.126	0	0	0	0	10.982	4.275	N
935-95-5	2,3,5,6-tetrachlorophenol	4.63	4.36	0.497	-0.151	0	0	0	0	13.085	4.225	N
4901-51-3	2,3,4,5-tetrachlorophenol	4.76	4.65	0.536	-0.178	0	0	0	0	13.085	4.170	N
2460-49-3	4,5-dichloroguaiacol	4.18	3.63	0.533	-0.106	0	0	0	0	5.920	4.170	N
2668-24-8	4,5,6-trichloroguaiacol	4.32	3.63	0.546	-0.161	0	0	0	0	7.488	4.110	N
57057-83-7	3,4,5-trichloroguaiacol	4.26	3.63	0.519	-0.118	0	0	0	0	7.488	4.180	N
2539-17-5	Tetrachloroguaiacol	4.44	3.63	0.529	-0.163	0	0	0	0	9.168	4.085	N
2539-26-6	Trichlorosyringol	3.90	3.38	0.480	-0.157	0	0	0	0	4.806	4.025	N
100-02-7	4-nitrophenol	3.88	3.49	0.516	-0.089	0	0	0	2	1.816	4.400	HPV
1689-84-5	Bromoxynil	4.50	4.58	0.583	-0.186	0	0	0	2	7.002	4.310	N
108-46-3	Resorcinol	3.27	3.12	0.478	-0.044	0	0	0	0	0.798	4.605	HPV
120-80-9	Catechol	3.39	3.12	0.472	-0.058	0	0	0	0	0.798	4.390	HPV

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
123-31-9	Hydroquinone	3.53	3.12	0.484	-0.069	0	0	0	0	0.798	4.235	HPV
615-67-8	Chlorohydroquinone	3.80	3.12	0.505	-0.075	0	0	0	0	2.293	4.160	N
95-88-5	4-chlororesorcinol	3.70	3.65	0.521	-0.053	0	0	0	0	2.293	4.415	N
2138-22-9	4-chlorocatechol	3.67	3.12	0.498	-0.083	0	0	0	0	2.293	4.265	N
3428-24-8	4,5-dichlorocatechol	4.00	3.65	0.525	-0.112	0	0	0	0	4.477	4.170	N
3978-67-4	3,4-dichlorocatechol	3.96	3.65	0.513	-0.099	0	0	0	0	4.477	4.200	N
13673-92-2	3,5-dichlorocatechol	3.84	3.12	0.496	-0.117	0	0	0	0	4.477	4.235	N
137-19-9	4,6-dichlororesorcinol	3.83	3.12	0.494	-0.084	0	0	0	0	4.477	4.325	N
56961-20-7	3,4,5-trichlorocatechol	4.11	3.63	0.528	-0.160	0	0	0	0	5.920	4.105	N
32139-72-3	3,4,6-trichlorocatechol	3.98	3.12	0.494	-0.163	0	0	0	0	5.920	4.115	N
1198-55-6	Tetrachlorocatechol	4.13	3.65	0.514	-0.228	0	0	0	0	7.488	4.050	N
87-87-6	Tetrachlorohydroquinone	4.20	3.38	0.486	-0.197	0	0	0	0	7.488	3.905	N
87-66-1	1,2,3-trihydroxybenzene	3.20	2.72	0.462	-0.089	0	0	0	0	0.106	4.430	N
99-55-8	2-Amino-4-nitrotoluene	4.59	3.56	0.463	-0.093	0	0	0	5	2.893	3.940	N
119-32-4	4-Amino-2-nitrotoluene	4.67	3.56	0.465	-0.081	0	0	0	5	2.893	3.880	N
603-83-8	2-Amino-6-nitrotoluene	4.77	3.56	0.452	-0.055	0	0	0	6	2.893	3.945	N
19406-51-0	4-Amino-2,6-dinitrotoluene	5.08	4.31	0.441	-0.100	0	0	0	8	3.055	3.795	N
35572-78-2	2-Amino-4,6-dinitrotoluene	4.90	4.31	0.429	-0.137	0	0	0	8	3.055	3.875	N
823-40-5	2,6-diaminotoluene	3.99	4.35	0.415	-0.126	0	0	0	6	1.553	4.450	HPV
6629-29-4	2,4-Diamino-6-nitrotoluene	4.81	4.31	0.441	-0.126	0	0	0	8	1.395	3.865	N
59229-75-3	2,6-Diamino-4-nitrotoluene	4.62	4.31	0.428	-0.186	0	0	0	8	1.395	3.910	N
56-75-7	Chloramphenicol	3.72	3.28	0.416	0.001	0	0	0	3	1.508	4.525	N
121-14-2	2,4-dinitrotoluene	4.35	3.52	0.455	-0.066	0	0	0	5	5.014	4.560	HPV
118-96-7	2,4,6-trinitrotoluene	4.65	4.31	0.422	-0.094	0	0	0	8	5.563	4.600	HPV
1194-65-6	2,6-dichlorobenzonitrile	4.57	3.81	0.528	-0.086	0	0	0	2	8.735	4.435	N
1897-45-6	Chlorothalonil	4.92	4.58	0.537	-0.277	0	0	0	4	9.814	4.075	HPV
34256-82-1	Acetochlor	5.10	4.76	0.386	0.252	0	0	0	5	10.127	4.715	HPV
23184-66-9	Butachlor	5.95	6.42	0.322	0.539	0	0	0	5	15.308	4.695	N
51218-49-6	Pretilachlor	5.54	4.67	0.384	0.455	0	0	0	4	12.296	4.730	N
15545-48-9	Chlorotoluron	4.83	3.84	0.461	0.044	0	0	0	5	5.654	4.320	HPV
23564-05-8	Thiophanate methyl	5.58	4.51	0.415	-0.189	0	0	1	10	0.026	4.050	N
57-67-0	Sulfaguanidine	2.84	3.83	0.478	-0.149	0	2	0	4	0.049	4.455	N
73231-34-2	Florfenicol	4.67	3.88	0.423	0.172	0	0	1	1	2.881	4.780	N
64249-01-0	Anilofos	5.20	3.88	0.386	0.131	0	0	1	2	5.292	4.320	N
22224-92-6	Fenamiphos	5.62	4.70	0.403	0.335	0	0	0	2	14.039	4.240	HPV
69377-81-7	Fluroxypyr	4.12	3.52	0.512	-0.124	0	0	0	4	1.682	4.355	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
21725-46-2	Cyanazine	6.01	5.35	0.432	0.284	0	0	0	13	4.124	4.615	N
834-12-8	Ametryn	7.30	6.21	0.436	0.236	0	0	1	13	6.707	4.410	HPV
7287-19-6	Prometryn	7.47	6.21	0.405	0.301	0	0	1	13	8.371	4.415	N
59-87-0	Nitrofurazone	3.98	3.83	0.529	-0.174	1	1	0	4	0.030	4.005	N
34014-18-1	Tebuthiuron	6.37	6.21	0.417	0.187	0	0	1	10	1.542	4.160	N
119-12-0	Pyridaphenthion	7.10	5.72	0.414	0.181	2	0	1	6	8.355	3.880	N
24096-53-5	Dimethachlon	4.23	2.87	0.466	0.017	0	0	0	2	5.997	4.520	N
36734-19-7	Iprodione	5.01	3.52	0.463	0.125	0	0	0	7	4.034	4.475	HPV
39807-15-3	Oxadiargyl	6.59	6.37	0.466	0.079	0	0	0	8	15.579	3.990	N
51338-27-3	Diclofop methyl	4.66	4.05	0.446	-0.055	0	0	0	0	13.399	4.230	HPV
40843-25-2	Diclofop P	4.45	4.49	0.471	-0.124	0	0	0	0	11.716	4.235	N
40843-73-0	4-(2,4-dichlorophenoxy)-phenol	4.74	4.65	0.515	-0.115	0	0	0	0	12.856	4.215	HPV
42874-03-3	Oxyfluorfen	5.18	4.69	0.421	-0.111	0	0	0	2	16.979	4.145	N
68359-37-5	Beta-cyfluthrin	5.52	5.35	0.350	0.141	1	0	0	1	18.311	4.300	N
54910-89-3	Fluoxetine	5.29	4.35	0.433	0.097	0	0	0	1	17.245	4.425	N
22071-15-4	Ketoprofen	4.41	3.71	0.405	0.116	0	0	0	0	11.384	4.550	N
85-68-7	Butylbenzyl phthalate	5.19	5.00	0.375	0.240	0	0	0	0	16.739	4.390	HPV
71626-11-4	Benalaxyl	5.09	4.76	0.365	0.126	0	0	0	5	13.285	4.720	N
126833-17-8	Fenhexamid	5.85	4.50	0.404	0.260	0	0	0	5	11.818	4.060	N
72619-32-0	Haloxyfop R	4.92	3.91	0.438	-0.025	0	0	0	3	11.626	4.285	N
83066-88-0	Fluazifop P	4.63	3.91	0.449	0.001	0	0	0	3	8.611	4.395	N
83055-99-6	Bensulfuron-methyl	5.95	4.51	0.371	0.006	0	0	1	10	3.119	4.250	N
90982-32-4	Chlorimuron-ethyl	6.21	4.51	0.410	0.131	0	0	1	10	2.815	4.355	N
111991-09-4	Nicosulfuron	7.06	6.76	0.418	0.257	0	0	1	14	0.155	4.035	N
136849-15-5	Cyclosulfamuron	6.94	6.26	0.364	0.047	0	0	1	14	5.834	4.160	N
74223-64-6	Metsulfuron-methyl	6.24	4.51	0.417	0.059	0	0	1	12	2.778	4.560	N
106040-48-6	Tribenuron	5.92	5.35	0.441	0.084	0	0	0	15	2.778	4.490	N
111353-84-5	Ethametsulfuron	6.17	4.43	0.454	0.067	0	0	0	16	1.838	4.255	N
79319-85-0	Bismerthiazol	5.87	6.32	0.534	-0.121	0	0	1	6	1.040	3.730	N
93697-74-6	Pyrazosulfuron ethyl	6.10	4.51	0.415	0.051	0	0	1	12	0.516	4.420	N
84087-01-4	Quinclorac	4.87	3.94	0.482	-0.095	0	0	0	4	7.609	4.065	N
52316-55-9	Carbendazim	4.86	3.24	0.505	-0.142	0	0	0	9	1.106	4.215	N
17804-35-2	Benomyl	6.26	5.35	0.420	0.173	0	0	0	15	2.760	4.180	N
18691-97-9	Methabenzthiazuron	6.12	5.28	0.455	-0.027	0	0	1	9	3.016	4.110	HPV
25059-80-7	Benazolin ethyl	4.46	3.58	0.426	0.147	0	0	0	2	4.714	4.195	N
260-94-6	Acridine	5.28	4.51	0.455	-0.094	0	0	0	5	9.309	3.815	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
59-40-5	Sulfaquinoxaline	5.64	5.12	0.439	-0.218	0	0	1	9	2.049	4.070	N
94051-08-8	Quizalofop P	4.76	4.49	0.442	-0.169	0	0	0	5	7.616	4.035	N
73250-68-7	Mefenacet	5.02	3.43	0.435	0.003	0	0	0	5	8.834	4.255	HPV
95617-09-7	Fenoxaprop	4.74	4.20	0.473	-0.101	0	0	0	3	9.023	4.165	N
98967-40-9	Flumetsulam	6.78	5.28	0.482	0.035	0	0	1	8	5.706	3.675	N
139-91-3	Furaltadone	5.24	4.26	0.444	0.150	1	0	0	7	0.260	3.980	N
87818-31-3	Cinmethylin	4.88	4.21	0.361	0.525	0	0	0	0	11.335	4.745	N
125401-75-4	Bispyribac	5.82	4.87	0.400	-0.025	0	0	0	12	8.882	4.355	N
564-25-0	Deoxytetracycline	4.39	5.33	0.348	0.255	0	1	0	5	1.922	3.860	N
57-62-5	Chlorotetracycline	4.70	5.33	0.359	0.403	0	1	0	5	0.816	3.730	N
100986-85-4	Levofloxacin	5.23	4.26	0.399	0.142	1	0	0	6	1.314	3.705	N
41083-11-8	Azocyclotin	9.35	6.42	0.323	1.201	0	0	0	9	30.509	4.650	N
76-87-9	Fentin hydroxide	4.96	4.68	0.405	-0.139	0	0	0	0	24.520	4.870	N
13121-70-5	Cyhexatin	6.61	4.36	0.334	1.007	0	0	0	0	21.129	4.835	N
55268-75-2	Cefuroxime	4.24	5.72	0.411	-0.027	0	2	1	5	0.744	4.150	N
26787-78-0	Amoxicillin	4.98	5.72	0.404	0.046	0	0	1	5	0.338	4.430	N
15686-71-2	Cephalexin	5.48	5.72	0.446	0.131	0	0	1	5	0.964	4.260	N
2022-85-7	5-fluorocytosine	3.70	3.44	0.486	-0.058	1	1	0	4	0.029	4.480	N
16110-51-3	Cromolyn	4.07	4.77	0.366	0.116	2	0	0	0	0.002	3.980	N
58-08-2	Caffeine	4.80	4.31	0.428	0.027	0	0	0	9	0.542	4.290	HPV
103-90-2	Acetaminophen	3.98	3.70	0.489	-0.022	0	0	0	2	1.123	4.220	HPV
73-22-3	L-tryptophan	4.08	3.67	0.436	-0.051	0	0	0	4	1.605	4.230	N
59-05-2	Methotrexate	4.93	4.04	0.374	0.005	0	3	0	16	1.719	3.620	N
51-52-5	Propylthiouracil	5.62	5.38	0.501	0.100	1	0	1	3	0.690	4.140	N
60-80-0	Antipyrine	5.00	4.26	0.433	0.006	1	0	0	5	5.317	4.140	HPV
87-08-1	Phenoxymethylpenicillinic acid	4.94	5.38	0.453	0.136	0	0	1	2	1.768	4.530	N
64544-07-6	Cefuroxime axetil	4.31	3.88	0.405	0.130	0	2	1	4	0.179	4.160	N
33419-42-0	Etoposide	3.44	4.27	0.338	0.117	0	0	0	0	0.018	4.060	N
51481-61-9	Cimetidine	4.78	4.31	0.408	0.093	0	0	0	9	0.674	4.410	N
94-20-2	Chlorpropamide	5.28	5.72	0.421	0.125	0	0	1	4	2.033	4.340	N
3930-20-9	Sotalol	5.27	3.88	0.438	0.290	0	0	1	3	0.503	4.450	N
58-93-5	Hydrochlorothiazide	3.37	3.83	0.448	-0.002	0	2	0	6	0.299	4.420	HPV
1156-19-0	Tolazamide	5.75	3.88	0.407	0.323	0	0	1	6	2.632	4.620	HPV
50-23-7	Hydrocortisone	4.72	3.39	0.338	0.748	1	0	0	0	2.619	4.610	N
42200-33-9	Nadolol	4.19	3.95	0.377	0.457	0	0	0	1	1.844	4.540	N
50-24-8	Prednisolone	5.09	4.18	0.341	0.734	3	0	0	0	2.334	4.690	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
51-34-3	Scopolamine	4.10	4.27	0.374	0.377	0	0	0	1	2.015	4.470	N
26839-75-8	Timolol	6.00	3.88	0.334	0.507	0	0	1	4	1.496	3.890	N
37350-58-6	Metoprolol	4.41	3.54	0.451	0.362	0	0	0	1	2.732	4.470	N
1091-85-6	Dansylglycine	5.57	5.72	0.413	0.092	0	0	1	5	1.724	3.990	N
137-58-6	Lidocaine	4.74	4.63	0.373	0.301	0	0	0	4	6.351	4.570	N
83-43-2	Methylprednisolone	5.33	4.18	0.327	0.843	3	0	0	0	3.051	4.680	N
64-77-7	Tolbutamide	5.29	3.88	0.400	0.236	0	0	1	4	2.070	4.510	N
526-08-9	Sulfaphenazole	5.57	5.12	0.419	-0.289	0	0	1	10	3.879	4.320	N
37517-30-9	Acebutolol	5.14	3.44	0.434	0.380	0	0	0	4	2.526	4.040	N
59-46-1	Procaine	4.26	3.57	0.429	0.031	0	0	0	2	4.506	4.150	N
63590-64-7	Terazosin	6.07	3.69	0.403	0.254	0	1	0	15	1.956	3.870	N
6452-71-7	Oxprenolol	4.45	3.39	0.394	0.366	1	0	0	1	3.351	4.550	N
84057-84-1	Lamotrigine	4.22	4.04	0.498	-0.032	0	3	0	8	7.772	4.230	N
4205-90-7	Clonidine	5.08	4.49	0.459	-0.050	0	0	0	6	7.074	4.130	N
13523-86-9	Pindolol	4.16	3.66	0.384	0.130	0	0	0	3	1.707	4.170	N
54-31-9	Furosemide	3.18	3.83	0.441	-0.143	0	2	0	6	0.189	4.270	HPV
66357-35-5	Ranitidine	4.65	3.88	0.344	-0.047	1	0	1	3	0.434	4.290	N
7689-03-4	Camptothecin	5.89	4.53	0.431	0.152	1	0	0	8	4.638	3.840	N
34841-39-9	Bupropion	5.29	4.83	0.406	0.268	0	0	0	1	10.355	3.850	N
103628-46-2	Sumatriptan	5.43	3.88	0.377	0.171	0	0	1	5	0.800	4.090	N
81-81-2	Warfarin	4.71	4.14	0.396	0.218	0	0	0	0	10.264	4.230	N
28395-03-1	Bumetanide	5.03	5.72	0.392	0.154	0	2	1	5	3.108	3.790	N
129-20-4	Oxyphenbutazone	6.12	4.69	0.390	0.105	0	0	0	10	10.110	4.010	N
87848-99-5	Acrivastine	6.36	4.03	0.377	0.312	3	0	0	5	8.673	3.920	N
57-41-0	Phenytoin	3.98	3.83	0.410	-0.083	0	0	0	6	3.221	4.770	HPV
564-25-0	Doxycycline	4.14	2.69	0.358	0.254	0	1	0	5	0.766	4.080	N
13655-52-2	Alprenolol	4.70	3.39	0.399	0.362	1	0	0	1	5.619	4.530	N
19216-56-9	Prazosin	5.67	3.69	0.428	0.020	0	1	0	15	1.541	3.870	N
36894-69-6	Labetalol	4.61	2.69	0.384	0.267	0	1	0	4	7.151	4.240	N
50-33-9	Phenylbutazone	6.62	5.89	0.391	0.201	0	0	0	10	13.685	4.060	N
637-07-0	Clofibrate	4.34	3.65	0.413	0.162	0	0	0	0	8.589	4.430	N
94-24-6	Tetracaine	5.10	3.83	0.450	0.245	0	0	0	3	6.983	4.200	N
6990-06-3	Fusidic acid	7.07	4.03	0.310	1.189	1	0	0	0	19.489	4.680	N
303-81-1	Novobiocin	4.07	3.04	0.394	0.309	1	2	0	4	0.387	3.970	N
99614-02-5	Ondansetron	5.44	3.15	0.402	0.226	0	0	0	7	6.074	4.160	N
548-73-2	Droperidol	6.27	4.64	0.391	0.226	1	0	0	9	9.420	4.150	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
56-54-2	Quinidine	5.60	4.64	0.358	0.356	1	0	0	6	4.805	4.020	N
53-86-1	Indomethacin	5.52	4.50	0.392	0.048	0	0	0	5	11.006	3.800	N
130-95-0	Quinine	5.67	4.64	0.365	0.379	1	0	0	6	4.805	4.020	N
599-79-1	Sulfasalazine	6.83	4.51	0.442	-0.136	0	0	1	13	3.875	3.750	N
57-83-0	Progesterone	6.03	4.03	0.339	0.838	1	0	0	0	16.399	4.900	N
50-47-5	Desipramine	5.89	4.50	0.371	0.035	0	0	0	8	13.278	4.060	N
50-28-2	Estradiol	5.26	4.22	0.382	0.492	0	0	0	0	13.211	4.510	N
10238-21-8	Glibenclamide	6.29	6.17	0.311	0.279	0	0	1	8	6.350	4.240	HPV
58-22-0	Testosterone	5.65	4.03	0.351	0.757	1	0	0	0	13.659	4.890	N
50-49-7	Imipramine	6.54	5.71	0.371	0.270	0	0	0	8	15.025	4.030	N
65277-42-1	Ketoconazole	5.38	4.50	0.358	0.105	0	0	0	4	8.996	3.510	N
58-40-2	Promazine	6.78	5.72	0.397	0.111	0	0	1	6	10.643	3.720	N
84625-61-6	Itraconazole	9.36	6.62	0.388	0.231	1	0	0	15	26.290	3.440	N
146-54-3	Triflupromazine	7.58	6.17	0.387	0.164	0	0	1	6	16.930	3.570	N
50-53-3	Chlorpromazine	7.16	5.72	0.397	0.104	0	0	1	6	14.197	3.660	N
91161-71-6	Terbinafine	7.16	6.11	0.399	0.242	2	0	0	3	24.838	4.170	N
23593-75-1	Clotrimazole	6.41	5.62	0.358	-0.024	0	0	0	7	24.255	4.380	N
3332-27-2*	**N,N-Dimethyltetradecylamine N-oxide**	**5.52**	**3.23**	**0.464**	**1.164**	**0**	**0**	**0**	**1**	**2.779**	**5.100**	**HPV**
116-37-0	1,1'-Isopropylidenebis(p-phenyleneoxy) Dipropan-2-ol	4.45	3.52	0.403	0.192	0	0	0	0	9.456	4.435	N
10222-01-2	2,2-Dibromo-2-cyanoacetamide	3.67	3.14	0.571	0.087	0	1	0	2	0.001	4.560	N
6021-61-0	Disperse red 54	6.36	4.87	0.421	-0.084	0	0	0	14	6.366	3.665	N
50-29-3	Dichlorodiphenyltrichloroethane (DDT)	7.17	4.84	0.449	0.074	0	0	0	0	38.020	4.445	HPV
309-00-2	Aldrin	7.15	4.86	0.427	0.381	2	0	0	0	28.743	4.570	N
36355-01-8	Hexabromobiphenyl	7.98	4.81	0.537	-0.409	0	0	0	0	51.159	4.245	N
101-14-4	2,2'-dichloro-4,4'-methylendianiline	5.37	3.98	0.439	-0.071	0	0	0	4	14.759	4.195	HPV
31508-00-6	1,2,4-trichloro-5-(3,4-dichlorophenyl)benzene	7.00	4.81	0.492	-0.191	0	0	0	0	38.839	4.290	N
208-96-8	Acenaphthylene	5.60	4.86	0.434	0.051	2	0	0	0	15.023	4.050	N
56-55-3	Benzo[a]anthracene	5.98	5.00	0.440	-0.228	0	0	0	0	26.674	3.745	N
53-70-3	Dibenzo[a,h]anthracene	6.64	5.38	0.434	-0.293	0	0	0	0	35.045	3.770	N
101-55-3	1-bromo-4-phenoxybenzene	5.16	4.36	0.532	-0.063	0	0	0	0	16.545	4.355	N
24017-47-8	Triazophos	6.56	5.38	0.454	0.237	0	0	1	5	8.874	4.160	N
1461-25-2	Tetra-n-butyltin	7.98	5.16	0.351	1.374	0	0	0	0	30.524	5.250	HPV
4640 01 1	Methyl triclosan	4.86	4.69	0.456	-0.122	0	0	0	0	16.737	4.280	N
112-18-5	N,N-dimethyldodecan-1-amine	6.71	4.63	0.452	0.973	0	0	0	1	24.151	5.745	HPV

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
124-19-6	Nonanal	5.30	3.96	0.481	0.600	1	0	0	0	11.277	5.250	HPV
526-73-8	1,2,3-trimethylbenzene	4.29	3.71	0.401	0.205	0	0	0	0	10.620	4.815	N
629-59-4	Tetradecane	6.58	4.99	0.456	1.051	0	0	0	0	35.181	7.345	HPV
1120-21-4	Undecane	5.20	4.99	0.455	0.820	0	0	0	0	26.162	7.400	HPV
84-65-1	9,10-Anthracenedione	4.32	3.84	0.448	0.034	0	0	0	0	8.365	4.270	HPV
86-74-8	Carbazole	4.78	4.94	0.453	-0.130	0	0	0	4	8.174	4.035	HPV
92-06-8	1,3-diphenylbenzene	5.69	5.14	0.440	-0.104	0	0	0	0	26.674	4.440	N
580-51-8	3-phenylphenol	4.15	4.49	0.465	-0.098	0	0	0	0	9.778	4.415	N
110-54-3	N-hexane	3.00	3.17	0.451	0.437	0	0	0	0	12.363	7.540	HPV
122-88-3	4-chlorophenoxyacetic acid	3.96	3.65	0.546	-0.015	0	0	0	0	3.179	4.400	N
80060-09-9	Diafenthiuron	8.51	6.17	0.358	0.496	0	0	1	6	24.116	3.970	N
59-30-3	Folic acid	5.16	3.69	0.390	-0.128	0	1	0	14	0.328	3.635	N
142469-14-5	Tritosulfuron	5.80	4.87	0.421	0.059	0	0	0	12	8.343	4.625	N
84030-86-4	Esbiothrin	5.85	4.55	0.386	0.676	2	0	0	0	11.553	4.625	N
5836 10 2	Chlorpropylate	5.18	4.36	0.386	0.240	0	0	0	0	17.344	4.525	N
78-34-2	Dioxathion	5.16	3.88	0.381	0.488	0	0	1	0	0.202	4.200	N
957-51-7	Diphenamid	4.64	3.71	0.389	0.180	0	0	0	2	11.191	4.705	N
2540-82-1	Formothion	3.76	4.94	0.409	0.099	1	0	0	0	0.733	4.315	N
961-22-8	Azinphosmethyl oxon	5.67	5.28	0.466	-0.082	0	0	1	8	0.920	4.150	N
16655-82-6	3-hydroxycarbofuran	3.70	3.43	0.426	0.099	0	0	0	1	0.451	4.330	N
3739-38-6	3-Phenoxybenzoic acid	4.23	4.49	0.454	-0.099	0	0	0	0	8.705	4.125	N
107-49-3	Tetraethyl pyrophosphate	2.88	2.23	0.366	0.210	0	0	0	0	4.203	5.695	N
7421-93-4	Endrin aldehyde	5.89	4.25	0.432	0.527	1	0	0	0	19.817	5.160	N
31972-43-7	Fenamiphos sulfoxide	4.95	4.14	0.389	0.262	0	0	0	2	8.256	4.125	N
2581-34-2	3-Methyl-4-nitrophenol	3.97	3.52	0.478	-0.007	0	0	0	3	1.291	4.440	N
3761-41-9	Fenthion sulfoxide	4.85	3.88	0.424	0.055	0	0	1	0	2.555	4.020	N
3761-42-0	Fenthion sulfone	4.31	3.88	0.425	0.033	0	0	1	0	2.373	4.660	N
87237-48-7	Haloxyfop-2-ethoxyethyl	4.97	3.72	0.398	0.003	0	0	0	3	10.980	4.035	N
2635 10 1	Methiocarb sulfoxide	4.12	3.16	0.463	0.068	0	0	0	1	2.202	4.095	N
2179-25-1	Methiocarb sulfone	3.75	2.87	0.464	0.127	0	0	0	1	2.033	4.705	N
2588 03 6	Phorate sulfoxide	4.87	3.88	0.401	0.216	0	0	1	0	0.142	3.990	N
2588 04 7	Phorate sulfone	4.84	3.88	0.430	0.289	0	0	1	0	0.081	4.335	N
1942-71-8	2-(4-tert-butylphenoxy)cyclohexanol	4.84	3.98	0.387	0.424	0	0	0	0	9.947	4.490	N
27304-13-8	Oxychlordane	5.89	4.66	0.413	0.434	0	0	0	0	19.817	4.535	N
53380-22-6	Ethiofencarb sulfoxide	4.74	5.72	0.417	-0.032	0	0	1	1	1.479	3.970	N
53380-23-7	Ethiofencarb sulfone	4.51	3.88	0.419	0.120	0	0	1	1	1.341	4.645	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
311-45-5	Ethyl paraoxon	4.05	3.28	0.423	0.039	0	0	0	2	4.502	4.420	N
40020-01-7	Pyridafol	5.29	3.27	0.516	0.056	1	0	0	5	4.349	4.140	N
2703-37-9	Thiometon sulfoxide	4.53	3.68	0.479	-0.002	0	0	1	0	0.001	4.220	N
20301-63-7	Thioometon sulfone	4.40	3.68	0.485	0.029	0	0	1	0	0.005	4.500	N
95-69-2	4-Chloro-2-methylaniline	4.45	4.00	0.469	-0.052	0	0	0	3	6.055	4.270	N
140-38-5	(4-chlorophenyl)urea	3.67	3.14	0.536	-0.128	0	1	0	3	2.148	4.330	N
61898-95-1	Methyl-3-(2,2-dichlorovinyl)-2,2-dimethyl-(1-cyclopropane) carboxylate	4.23	4.70	0.481	0.172	1	0	0	0	6.114	4.905	N
1713-15-1	2,4-D-1-isobutyl ester	4.74	3.69	0.488	0.073	0	0	0	0	11.987	4.460	N
62610-77-9	Methacrifos	4.52	3.68	0.482	0.080	1	0	1	0	0.096	4.770	N
2227-13-6	Tetrasul	7.17	4.81	0.504	-0.075	0	0	0	0	35.867	4.040	N
950-10-7	Mephospholan	5.36	5.72	0.433	0.131	0	0	1	4	2.375	4.345	N
1214-39-7	6-Benzyladenine	5.54	4.69	0.432	-0.018	0	0	0	10	6.714	4.215	N
120923-37-7	Amidosulfuron	5.24	4.31	0.449	0.155	0	0	0	11	0.054	4.465	N
120162-55-2	Azimsulfuron	6.21	4.43	0.417	-0.009	0	0	0	19	1.991	4.450	N
120-23-0	2-Naphthyloxyacetic acid	4.06	3.38	0.490	-0.090	0	0	0	0	4.939	4.050	N
41483-43-6	Bupirimate	7.42	6.17	0.397	0.543	0	0	1	9	6.626	4.045	N
55285-14-8	Carbosulfan	7.40	6.17	0.321	0.716	0	0	1	7	9.357	4.125	N
1134-23-2	Cycloate	5.63	3.88	0.415	0.362	0	0	1	3	4.727	4.595	HPV
13684-56-5	Desmedipham	4.63	3.98	0.472	-0.072	0	0	0	6	4.354	4.375	N
3347-22-6	Dithianon	6.07	5.49	0.502	0.124	0	0	1	2	1.735	3.240	N
126801-58-9	Ethoxysulfuron	6.54	4.51	0.415	0.088	0	0	1	10	3.987	3.990	N
61213-25-0	Fluorochloridone	5.35	3.69	0.434	0.127	0	0	0	3	13.058	4.280	N
77-06-5	Gibberellic acid	4.38	4.55	0.348	0.525	2	0	0	0	3.723	4.995	N
10004-44-1	Hymexazol	3.60	4.29	0.516	-0.037	1	0	0	1	0.006	4.775	N
81405-85-8	Imazamethabenz-methyl	5.53	4.63	0.376	0.251	0	0	0	8	7.628	4.380	N
140923-17-7	Iprovalicarb	5.10	4.63	0.349	0.479	0	0	0	6	5.613	4.740	N
123-33-1	Maleic hydrazide	4.32	5.58	0.471	-0.057	2	0	0	2	0.093	4.095	N
133408-50-1	Metominostrobin	4.41	3.90	0.394	0.112	0	0	0	3	6.015	4.395	N
2310-17-0	Phosalone	4.70	5.72	0.429	-0.139	0	0	1	2	2.740	4.160	N
90717-03-6	Quinmerac	5.03	3.52	0.485	0.001	0	0	0	5	6.202	4.125	N
111872-58-3	Halfenprox	6.15	5.06	0.353	0.010	0	0	0	0	25.763	3.615	N
90035-08-8	Flocoumafen	8.53	5.33	0.410	0.437	0	0	0	0	42.453	3.960	N
65731-84-2	Beta cypermethrin	5.01	5.35	0.346	0.052	1	0	0	1	15.264	4.340	N
56073-10-0	Brodifacoum	8.12	5.38	0.401	0.143	0	0	0	0	44.742	4.020	N
1469-48-3	Cis-1,2,3,6-Tetrahydrophthalimide	4.07	4.77	0.426	0.222	2	0	0	2	0.395	4.965	N

Table B.2. Continued.

CAS	Name	Pred pEC_{50} (GM)	Pred pEC_{50} (CPANN)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05[C-S]	F03[C-N]	MLOGP2	Hardness	HPV Status[a]
6515-38-4	3,5,6-trichloro-2-pyridinol	4.60	3.39	0.524	-0.071	1	0	0	1	4.536	4.005	N
1031-07-8	Endosulfan sulfate	5.97	5.51	0.443	0.341	0	0	1	0	11.757	4.540	N
120068-36-2	Fipronil sulfone	5.64	5.28	0.456	-0.217	0	1	1	8	6.821	3.990	N
120067-83-6	Fipronil sulfide	6.28	5.28	0.461	-0.202	0	1	1	8	12.367	3.920	N
1689-83-4	Ioxynil	4.72	4.58	0.600	-0.323	0	0	0	2	8.629	3.965	N
1646-87-3	Aldicarb-sulfoxide	5.22	5.38	0.474	0.119	1	0	1	2	0.009	4.305	N
1646-88-4	Aldicarb-sulfone	4.80	3.68	0.469	0.198	1	0	1	2	0.022	5.025	N
3032-40-4	Fluometuron desmethyl	4.61	3.52	0.496	-0.020	0	0	0	4	4.988	4.320	N
1570-64-5	2-methyl-4-chlorophenol	4.06	3.26	0.476	0.012	0	0	0	0	6.055	4.385	HPV
94-80-4	2,4-D-1 -butyl ester	5.04	3.38	0.524	0.164	0	0	0	0	11.987	4.460	N
789-02-6	o,p'-DDT	7.33	4.84	0.443	0.174	0	0	0	0	38.020	4.455	N
67564-91-4	Fenpropimorph	6.08	6.42	0.354	0.609	0	0	0	4	14.642	4.570	HPV
319-84-6	HCH-alpha	4.25	4.68	0.419	-0.030	0	0	0	0	16.744	5.135	N
319-85-7	HCH-delta	3.95	2.86	0.417	-0.098	0	0	0	0	16.744	5.355	N
103055-07-8	Lufenuron	6.52	5.31	0.506	-0.037	0	0	0	6	18.643	3.970	N
119168-77-3	Tebufenpyrad	5.77	4.76	0.398	0.332	0	0	0	6	11.907	4.525	N
16484-77-8	Mecoprop-P	3.83	3.43	0.439	0.042	0	0	0	0	5.654	4.540	HPV
1746-81-2	Monolinuron	4.59	4.00	0.468	0.033	0	0	0	4	4.370	4.275	N
52888-80-9	Prosulfocarb	6.05	3.88	0.389	0.327	0	0	1	3	9.536	4.455	N
52315-07-8	Zeta-cypermethrin	5.28	4.26	0.356	0.193	1	0	0	1	15.264	4.380	N
66841-25-6	Tralomethrin	6.25	5.06	0.399	0.104	0	0	0	1	22.983	3.765	N
563-12-2	Ethion	4.39	3.88	0.424	0.023	0	0	1	0	0.392	4.270	N
70124-77-5	Flucythrinate	5.06	4.69	0.350	0.076	0	0	0	1	18.029	4.395	N
52918-63-5	Deltamethrin	5.66	5.35	0.408	0.150	1	0	0	1	16.930	4.225	N
139968-49-3	Metaflumizone	6.87	5.20	0.413	-0.036	0	0	0	10	20.097	4.055	N
70630-17-0	Metalaxyl-M	4.06	3.66	0.374	0.138	0	0	0	4	3.636	4.710	N
108-62-3	Metaldehyde	2.45	2.59	0.371	0.314	0	0	0	0	0.963	6.125	HPV
422556-08-9	Pyroxsulam	7.19	6.21	0.410	0.075	0	0	1	13	6.392	3.980	N
87820-88-0	Tralkoxydim	5.79	3.44	0.397	0.668	0	0	0	4	7.486	4.380	N
43121-43-3	Triadimefon	5.10	3.57	0.392	0.098	0	0	0	4	11.052	4.280	HPV
55219-65-3	Triadimenol	4.86	3.57	0.385	0.023	0	0	0	4	11.607	4.450	HPV
2303-17-5	Triallate	5.15	3.81	0.430	0.404	0	0	0	1	9.069	4.315	HPV
52-68-6	Trichlorphon (Chlorphos)	3.40	3.13	0.489	0.149	0	0	0	0	0.239	4.905	N
80844-07-1	Etofenprox	5.76	4.89	0.333	0.199	0	0	0	0	22.757	4.125	N
102851-06-9	Tau-fluvalinate	6.67	6.42	0.339	0.266	0	0	0	5	23.140	4.150	N

*From this compound to the end: Chemicals with no ecotoxicological data (SU0303, 2015). [a]Production volume status according to OECD (2009). HPV: High production volume. N: Not HPV.

APPENDIX C: DETAILED RESULTS OF IMPACT OF GEOMETRY OPTIMIZATION

Table C.1. List of molecular descriptors affected by quantum chemical method (Pearson correlations between pairs of descriptors: $r < 1$).

Descriptor	Chemical meaning	Type*
G1	Gravitational index G1	Geometrical descriptors
G2	Gravitational index G2 (bond-restricted)	Geometrical descriptors
RGyr	Radius of gyration (mass weighted)	Geometrical descriptors
SPAN	Span R	Geometrical descriptors
SPAM	Average span R	Geometrical descriptors
MEcc	Molecular eccentricity	Geometrical descriptors
SPH	Spherosity	Geometrical descriptors
ASP	Asphericity	Geometrical descriptors
PJI3	3D Petitjean shape index	Geometrical descriptors
L/Bw	Length-to-breadth ratio by WHIM	Geometrical descriptors
HOMA	Harmonic Oscillator Model of Aromaticity index	Geometrical descriptors
CMBL	Conjugated maximum bond length	Geometrical descriptors
AROM	Aromaticity index	Geometrical descriptors
HOMT	HOMA total	Geometrical descriptors
DISPm	Displacement value / weighted by mass	Geometrical descriptors
QXXm	Quadrupole x-component value / weighted by mass	Geometrical descriptors
QYYm	Quadrupole y-component value / weighted by mass	Geometrical descriptors
QZZm	Quadrupole z-component value / weighted by mass	Geometrical descriptors
DISPv	Displacement value / weighted by van der Waals volume	Geometrical descriptors
QXXv	Quadrupole x-component value / weighted by van der Waals volume	Geometrical descriptors
QYYv	Quadrupole y-component value / weighted by van der Waals volume	Geometrical descriptors
QZZv	Quadrupole z-component value / weighted by van der Waals volume	Geometrical descriptors
DISPe	Displacement value / weighted by Sanderson electronegativity	Geometrical descriptors
QXXe	Quadrupole x-component value / weighted by Sanderson electronegativity	Geometrical descriptors
QYYe	Quadrupole y-component value / weighted by Sanderson electronegativity	Geometrical descriptors
QZZe	Quadrupole z-component value / weighted by Sanderson electronegativity	Geometrical descriptors
DISPp	Displacement value / weighted by polarizability	Geometrical descriptors
QXXp	Quadrupole x-component value / weighted by polarizability	Geometrical descriptors
QYYp	Quadrupole y-component value / weighted by polarizability	Geometrical descriptors
QZZp	Quadrupole z-component value / weighted by polarizability	Geometrical descriptors
DISPi	Displacement value / weighted by ionization potential	Geometrical descriptors
QXXi	Quadrupole x-component value / weighted by ionization potential	Geometrical descriptors
QYYi	Quadrupole y-component value / weighted by ionization potential	Geometrical descriptors
QZZi	Quadrupole z-component value / weighted by ionization potential	Geometrical descriptors
DISPs	Displacement value / weighted by I-state	Geometrical descriptors
QXXs	Quadrupole x-component value / weighted by I-state	Geometrical descriptors
QYYs	Quadrupole y-component value / weighted by I-state	Geometrical descriptors
QZZs	Quadrupole z-component value / weighted by I-state	Geometrical descriptors
Wi_G	Wiener-like index from geometrical matrix	3D matrix-based descriptors
WiA_G	Average Wiener-like index from geometrical matrix	3D matrix-based descriptors
AVS_G	Average vertex sum from geometrical matrix	3D matrix-based descriptors
H_G	Harary-like index from geometrical matrix	3D matrix-based descriptors
Chi_G	Randic-like index from geometrical matrix	3D matrix-based descriptors
ChiA_G	Average Randic-like index from geometrical matrix	3D matrix-based descriptors
J_G	Balaban-like index from geometrical matrix	3D matrix-based descriptors
HyWi_G	Hyper-Wiener-like index from geometrical matrix	3D matrix-based descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
SpAbs_G	Graph energy from geometrical matrix	3D matrix-based descriptors
SpPos_G	Spectral positive sum from geometrical matrix	3D matrix-based descriptors
SpPosA_G	Normalized spectral positive sum from geometrical matrix	3D matrix-based descriptors
SpPosLog_G	Logarithmic spectral positive sum from geometrical matrix	3D matrix-based descriptors
SpMax_G	Leading eigenvalue from geometrical matrix	3D matrix-based descriptors
SpMaxA_G	Normalized leading eigenvalue from geometrical matrix	3D matrix-based descriptors
SpDiam_G	Spectral diameter from geometrical matrix	3D matrix-based descriptors
SpAD_G	Spectral absolute deviation from geometrical matrix	3D matrix-based descriptors
SpMAD_G	Spectral mean absolute deviation from geometrical matrix	3D matrix-based descriptors
Ho_G	Hosoya-like index (Log function) from geometrical matrix	3D matrix-based descriptors
SM2_G	Spectral moment of order 2 from geometrical matrix	3D matrix-based descriptors
SM3_G	Spectral moment of order 3 from geometrical matrix	3D matrix-based descriptors
SM4_G	Spectral moment of order 4 from geometrical matrix	3D matrix-based descriptors
SM5_G	Spectral moment of order 5 from geometrical matrix	3D matrix-based descriptors
SM6_G	Spectral moment of order 6 from geometrical matrix	3D matrix-based descriptors
VE1_G	Coefficient sum of the last eigenvector from geometrical matrix	3D matrix-based descriptors
VE2_G	Average coefficient of the last eigenvector from geometrical matrix	3D matrix-based descriptors
VE3_G	Logarithmic coefficient sum of the last eigenvector from geometrical matrix	3D matrix-based descriptors
VR1_G	Randic-like eigenvector-based index from geometrical matrix	3D matrix-based descriptors
VR2_G	Normalized Randic-like eigenvector-based index from geometrical matrix	3D matrix-based descriptors
VR3_G	Logarithmic Randic-like eigenvector-based index from geometrical matrix	3D matrix-based descriptors
Wi_RG	Wiener-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
WiA_RG	Average Wiener-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
AVS_RG	Average vertex sum from reciprocal squared geometrical matrix	3D matrix-based descriptors
H_RG	Harary-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
Chi_RG	Randic-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
ChiA_RG	Average Randic-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
J_RG	Balaban-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
HyWi_RG	Hyper-Wiener-like index from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpAbs_RG	Graph energy from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpPos_RG	Spectral positive sum from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpPosA_RG	Normalized spectral positive sum from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpPosLog_RG	Logarithmic spectral positive sum from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpMax_RG	Leading eigenvalue from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpMaxA_RG	Normalized leading eigenvalue from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpDiam_RG	Spectral diameter from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpAD_RG	Spectral absolute deviation from reciprocal squared geometrical matrix	3D matrix-based descriptors
SpMAD_RG	Spectral mean absolute deviation from reciprocal squared geometrical matrix	3D matrix-based descriptors
Ho_RG	Hosoya-like index (Log function) from reciprocal squared geometrical matrix	3D matrix-based descriptors
EE_RG	Estrada-like index (Log function) from reciprocal squared geometrical matrix	3D matrix-based descriptors
SM2_RG	Spectral moment of order 2 from reciprocal squared geometrical matrix	3D matrix-based descriptors
SM3_RG	Spectral moment of order 3 from reciprocal squared geometrical matrix	3D matrix-based descriptors
SM4_RG	Spectral moment of order 4 from reciprocal squared geometrical matrix	3D matrix-based descriptors
SM5_RG	Spectral moment of order 5 from reciprocal squared geometrical matrix	3D matrix-based descriptors
SM6_RG	Spectral moment of order 6 from reciprocal squared geometrical matrix	3D matrix-based descriptors
VE1_RG	Coefficient sum of the last eigenvector from reciprocal squared geometrical matrix	3D matrix-based descriptors
VE2_RG	Average coefficient of the last eigenvector from reciprocal squared geometrical matrix	3D matrix-based descriptors
VE3_RG	Logarithmic coefficient sum of the last eigenvector from reciprocal squared geometrical matrix	3D matrix-based descriptors
VR1_RG	Randic-like eigenvector-based index from reciprocal squared geometrical matrix	3D matrix-based descriptors
VR2_RG	Normalized Randic-like eigenvector-based index from reciprocal squared geometrical matrix	3D matrix-based descriptors
VR3_RG	Logarithmic Randic-like eigenvector-based index from reciprocal squared geometrical matrix	3D matrix-based descriptors
Wi_G/D	Wiener-like index from distance/distance matrix	3D matrix-based descriptors
WiA_G/D	Average Wiener-like index from distance/distance matrix	3D matrix-based descriptors
AVS_G/D	Average vertex sum from distance/distance matrix	3D matrix-based descriptors
H_G/D	Harary-like index from distance/distance matrix	3D matrix-based descriptors
Chi_G/D	Randic-like index from distance/distance matrix	3D matrix-based descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
ChiA_G/D	Average Randic-like index from distance/distance matrix	3D matrix-based descriptors
J_G/D	Balaban-like index from distance/distance matrix	3D matrix-based descriptors
HyWi_G/D	Hyper-Wiener-like index from distance/distance matrix	3D matrix-based descriptors
SpAbs_G/D	Graph energy from distance/distance matrix	3D matrix-based descriptors
SpPos_G/D	Spectral positive sum from distance/distance matrix	3D matrix-based descriptors
SpPosA_G/D	Normalized spectral positive sum from distance/distance matrix	3D matrix-based descriptors
SpPosLog_G/D	Logarithmic spectral positive sum from distance/distance matrix	3D matrix-based descriptors
SpMax_G/D	Leading eigenvalue from distance/distance matrix	3D matrix-based descriptors
SpMaxA_G/D	Normalized leading eigenvalue from distance/distance matrix (folding degree index)	3D matrix-based descriptors
SpDiam_G/D	Spectral diameter from distance/distance matrix	3D matrix-based descriptors
SpAD_G/D	Spectral absolute deviation from distance/distance matrix	3D matrix-based descriptors
SpMAD_G/D	Spectral mean absolute deviation from distance/distance matrix	3D matrix-based descriptors
Ho_G/D	Hosoya-like index (Log function) from distance/distance matrix	3D matrix-based descriptors
EE_G/D	Estrada-like index (Log function) from distance/distance matrix	3D matrix-based descriptors
SM2_G/D	Spectral moment of order 2 from distance/distance matrix	3D matrix-based descriptors
SM3_G/D	Spectral moment of order 3 from distance/distance matrix	3D matrix-based descriptors
SM4_G/D	Spectral moment of order 4 from distance/distance matrix	3D matrix-based descriptors
SM5_G/D	Spectral moment of order 5 from distance/distance matrix	3D matrix-based descriptors
SM6_G/D	Spectral moment of order 6 from distance/distance matrix	3D matrix-based descriptors
VE1_G/D	Coefficient sum of the last eigenvector from distance/distance matrix	3D matrix-based descriptors
VE2_G/D	Average coefficient of the last eigenvector from distance/distance matrix	3D matrix-based descriptors
VE3_G/D	Logarithmic coefficient sum of the last eigenvector from distance/distance matrix	3D matrix-based descriptors
VR1_G/D	Randic-like eigenvector-based index from distance/distance matrix	3D matrix-based descriptors
VR2_G/D	Normalized Randic-like eigenvector-based index from distance/distance matrix	3D matrix-based descriptors
VR3_G/D	Logarithmic Randic-like eigenvector-based index from distance/distance matrix	3D matrix-based descriptors
TDB01u	3D Topological distance based descriptors - lag 1 unweighted	3D autocorrelations
TDB02u	3D Topological distance based descriptors - lag 2 unweighted	3D autocorrelations
TDB03u	3D Topological distance based descriptors - lag 3 unweighted	3D autocorrelations
TDB04u	3D Topological distance based descriptors - lag 4 unweighted	3D autocorrelations
TDB05u	3D Topological distance based descriptors - lag 5 unweighted	3D autocorrelations
TDB06u	3D Topological distance based descriptors - lag 6 unweighted	3D autocorrelations
TDB07u	3D Topological distance based descriptors - lag 7 unweighted	3D autocorrelations
TDB08u	3D Topological distance based descriptors - lag 8 unweighted	3D autocorrelations
TDB09u	3D Topological distance based descriptors - lag 9 unweighted	3D autocorrelations
TDB10u	3D Topological distance based descriptors - lag 10 unweighted	3D autocorrelations
TDB01m	3D Topological distance based descriptors - lag 1 weighted by mass	3D autocorrelations
TDB02m	3D Topological distance based descriptors - lag 2 weighted by mass	3D autocorrelations
TDB03m	3D Topological distance based descriptors - lag 3 weighted by mass	3D autocorrelations
TDB04m	3D Topological distance based descriptors - lag 4 weighted by mass	3D autocorrelations
TDB05m	3D Topological distance based descriptors - lag 5 weighted by mass	3D autocorrelations
TDB06m	3D Topological distance based descriptors - lag 6 weighted by mass	3D autocorrelations
TDB07m	3D Topological distance based descriptors - lag 7 weighted by mass	3D autocorrelations
TDB08m	3D Topological distance based descriptors - lag 8 weighted by mass	3D autocorrelations
TDB09m	3D Topological distance based descriptors - lag 9 weighted by mass	3D autocorrelations
TDB10m	3D Topological distance based descriptors - lag 10 weighted by mass	3D autocorrelations
TDB01v	3D Topological distance based descriptors - lag 1 weighted by van der Waals volume	3D autocorrelations
TDB02v	3D Topological distance based descriptors - lag 2 weighted by van der Waals volume	3D autocorrelations
TDB03v	3D Topological distance based descriptors - lag 3 weighted by van der Waals volume	3D autocorrelations
TDB04v	3D Topological distance based descriptors - lag 4 weighted by van der Waals volume	3D autocorrelations
TDB05v	3D Topological distance based descriptors - lag 5 weighted by van der Waals volume	3D autocorrelations
TDB06v	3D Topological distance based descriptors - lag 6 weighted by van der Waals volume	3D autocorrelations
TDB07v	3D Topological distance based descriptors - lag 7 weighted by van der Waals volume	3D autocorrelations
TDB08v	3D Topological distance based descriptors - lag 8 weighted by van der Waals volume	3D autocorrelations
TDB09v	3D Topological distance based descriptors - lag 9 weighted by van der Waals volume	3D autocorrelations
TDB10v	3D Topological distance based descriptors - lag 10 weighted by van der Waals volume	3D autocorrelations
TDB01e	3D Topological distance based descriptors - lag 1 weighted by Sanderson electronegativity	3D autocorrelations

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
TDB02e	3D Topological distance based descriptors - lag 2 weighted by Sanderson electronegativity	3D autocorrelations
TDB03e	3D Topological distance based descriptors - lag 3 weighted by Sanderson electronegativity	3D autocorrelations
TDB04e	3D Topological distance based descriptors - lag 4 weighted by Sanderson electronegativity	3D autocorrelations
TDB05e	3D Topological distance based descriptors - lag 5 weighted by Sanderson electronegativity	3D autocorrelations
TDB06e	3D Topological distance based descriptors - lag 6 weighted by Sanderson electronegativity	3D autocorrelations
TDB07e	3D Topological distance based descriptors - lag 7 weighted by Sanderson electronegativity	3D autocorrelations
TDB08e	3D Topological distance based descriptors - lag 8 weighted by Sanderson electronegativity	3D autocorrelations
TDB09e	3D Topological distance based descriptors - lag 9 weighted by Sanderson electronegativity	3D autocorrelations
TDB10e	3D Topological distance based descriptors - lag 10 weighted by Sanderson electronegativity	3D autocorrelations
TDB01p	3D Topological distance based descriptors - lag 1 weighted by polarizability	3D autocorrelations
TDB02p	3D Topological distance based descriptors - lag 2 weighted by polarizability	3D autocorrelations
TDB03p	3D Topological distance based descriptors - lag 3 weighted by polarizability	3D autocorrelations
TDB04p	3D Topological distance based descriptors - lag 4 weighted by polarizability	3D autocorrelations
TDB05p	3D Topological distance based descriptors - lag 5 weighted by polarizability	3D autocorrelations
TDB06p	3D Topological distance based descriptors - lag 6 weighted by polarizability	3D autocorrelations
TDB07p	3D Topological distance based descriptors - lag 7 weighted by polarizability	3D autocorrelations
TDB08p	3D Topological distance based descriptors - lag 8 weighted by polarizability	3D autocorrelations
TDB09p	3D Topological distance based descriptors - lag 9 weighted by polarizability	3D autocorrelations
TDB10p	3D Topological distance based descriptors - lag 10 weighted by polarizability	3D autocorrelations
TDB01i	3D Topological distance based descriptors - lag 1 weighted by ionization potential	3D autocorrelations
TDB02i	3D Topological distance based descriptors - lag 2 weighted by ionization potential	3D autocorrelations
TDB03i	3D Topological distance based descriptors - lag 3 weighted by ionization potential	3D autocorrelations
TDB04i	3D Topological distance based descriptors - lag 4 weighted by ionization potential	3D autocorrelations
TDB05i	3D Topological distance based descriptors - lag 5 weighted by ionization potential	3D autocorrelations
TDB06i	3D Topological distance based descriptors - lag 6 weighted by ionization potential	3D autocorrelations
TDB07i	3D Topological distance based descriptors - lag 7 weighted by ionization potential	3D autocorrelations
TDB08i	3D Topological distance based descriptors - lag 8 weighted by ionization potential	3D autocorrelations
TDB09i	3D Topological distance based descriptors - lag 9 weighted by ionization potential	3D autocorrelations
TDB10i	3D Topological distance based descriptors - lag 10 weighted by ionization potential	3D autocorrelations
TDB01s	3D Topological distance based descriptors - lag 1 weighted by I-state	3D autocorrelations
TDB02s	3D Topological distance based descriptors - lag 2 weighted by I-state	3D autocorrelations
TDB03s	3D Topological distance based descriptors - lag 3 weighted by I-state	3D autocorrelations
TDB04s	3D Topological distance based descriptors - lag 4 weighted by I-state	3D autocorrelations
TDB05s	3D Topological distance based descriptors - lag 5 weighted by I-state	3D autocorrelations
TDB06s	3D Topological distance based descriptors - lag 6 weighted by I-state	3D autocorrelations
TDB07s	3D Topological distance based descriptors - lag 7 weighted by I-state	3D autocorrelations
TDB08s	3D Topological distance based descriptors - lag 8 weighted by I-state	3D autocorrelations
TDB09s	3D Topological distance based descriptors - lag 9 weighted by I-state	3D autocorrelations
TDB10s	3D Topological distance based descriptors - lag 10 weighted by I-state	3D autocorrelations
TDB01r	3D Topological distance based descriptors - lag 1 weighted by covalent radius	3D autocorrelations
TDB02r	3D Topological distance based descriptors - lag 2 weighted by covalent radius	3D autocorrelations
TDB03r	3D Topological distance based descriptors - lag 3 weighted by covalent radius	3D autocorrelations
TDB04r	3D Topological distance based descriptors - lag 4 weighted by covalent radius	3D autocorrelations
TDB05r	3D Topological distance based descriptors - lag 5 weighted by covalent radius	3D autocorrelations
TDB06r	3D Topological distance based descriptors - lag 6 weighted by covalent radius	3D autocorrelations
TDB07r	3D Topological distance based descriptors - lag 7 weighted by covalent radius	3D autocorrelations
TDB08r	3D Topological distance based descriptors - lag 8 weighted by covalent radius	3D autocorrelations
TDB09r	3D Topological distance based descriptors - lag 9 weighted by covalent radius	3D autocorrelations
TDB10r	3D Topological distance based descriptors - lag 10 weighted by covalent radius	3D autocorrelations
RDF010u	Radial Distribution Function - 010 / unweighted	RDF descriptors
RDF015u	Radial Distribution Function - 015 / unweighted	RDF descriptors
RDF020u	Radial Distribution Function - 020 / unweighted	RDF descriptors
RDF025u	Radial Distribution Function - 025 / unweighted	RDF descriptors
RDF030u	Radial Distribution Function - 030 / unweighted	RDF descriptors
RDF035u	Radial Distribution Function - 035 / unweighted	RDF descriptors
RDF040u	Radial Distribution Function - 040 / unweighted	RDF descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
RDF045u	Radial Distribution Function - 045 / unweighted	RDF descriptors
RDF050u	Radial Distribution Function - 050 / unweighted	RDF descriptors
RDF055u	Radial Distribution Function - 055 / unweighted	RDF descriptors
RDF060u	Radial Distribution Function - 060 / unweighted	RDF descriptors
RDF065u	Radial Distribution Function - 065 / unweighted	RDF descriptors
RDF070u	Radial Distribution Function - 070 / unweighted	RDF descriptors
RDF075u	Radial Distribution Function - 075 / unweighted	RDF descriptors
RDF080u	Radial Distribution Function - 080 / unweighted	RDF descriptors
RDF085u	Radial Distribution Function - 085 / unweighted	RDF descriptors
RDF090u	Radial Distribution Function - 090 / unweighted	RDF descriptors
RDF095u	Radial Distribution Function - 095 / unweighted	RDF descriptors
RDF100u	Radial Distribution Function - 100 / unweighted	RDF descriptors
RDF105u	Radial Distribution Function - 105 / unweighted	RDF descriptors
RDF110u	Radial Distribution Function - 110 / unweighted	RDF descriptors
RDF115u	Radial Distribution Function - 115 / unweighted	RDF descriptors
RDF120u	Radial Distribution Function - 120 / unweighted	RDF descriptors
RDF125u	Radial Distribution Function - 125 / unweighted	RDF descriptors
RDF130u	Radial Distribution Function - 130 / unweighted	RDF descriptors
RDF135u	Radial Distribution Function - 135 / unweighted	RDF descriptors
RDF140u	Radial Distribution Function - 140 / unweighted	RDF descriptors
RDF145u	Radial Distribution Function - 145 / unweighted	RDF descriptors
RDF150u	Radial Distribution Function - 150 / unweighted	RDF descriptors
RDF155u	Radial Distribution Function - 155 / unweighted	RDF descriptors
RDF010m	Radial Distribution Function - 010 / weighted by mass	RDF descriptors
RDF015m	Radial Distribution Function - 015 / weighted by mass	RDF descriptors
RDF020m	Radial Distribution Function - 020 / weighted by mass	RDF descriptors
RDF025m	Radial Distribution Function - 025 / weighted by mass	RDF descriptors
RDF030m	Radial Distribution Function - 030 / weighted by mass	RDF descriptors
RDF035m	Radial Distribution Function - 035 / weighted by mass	RDF descriptors
RDF040m	Radial Distribution Function - 040 / weighted by mass	RDF descriptors
RDF045m	Radial Distribution Function - 045 / weighted by mass	RDF descriptors
RDF050m	Radial Distribution Function - 050 / weighted by mass	RDF descriptors
RDF055m	Radial Distribution Function - 055 / weighted by mass	RDF descriptors
RDF060m	Radial Distribution Function - 060 / weighted by mass	RDF descriptors
RDF065m	Radial Distribution Function - 065 / weighted by mass	RDF descriptors
RDF070m	Radial Distribution Function - 070 / weighted by mass	RDF descriptors
RDF075m	Radial Distribution Function - 075 / weighted by mass	RDF descriptors
RDF080m	Radial Distribution Function - 080 / weighted by mass	RDF descriptors
RDF085m	Radial Distribution Function - 085 / weighted by mass	RDF descriptors
RDF090m	Radial Distribution Function - 090 / weighted by mass	RDF descriptors
RDF095m	Radial Distribution Function - 095 / weighted by mass	RDF descriptors
RDF100m	Radial Distribution Function - 100 / weighted by mass	RDF descriptors
RDF105m	Radial Distribution Function - 105 / weighted by mass	RDF descriptors
RDF110m	Radial Distribution Function - 110 / weighted by mass	RDF descriptors
RDF115m	Radial Distribution Function - 115 / weighted by mass	RDF descriptors
RDF120m	Radial Distribution Function - 120 / weighted by mass	RDF descriptors
RDF125m	Radial Distribution Function - 125 / weighted by mass	RDF descriptors
RDF130m	Radial Distribution Function - 130 / weighted by mass	RDF descriptors
RDF135m	Radial Distribution Function - 135 / weighted by mass	RDF descriptors
RDF140m	Radial Distribution Function - 140 / weighted by mass	RDF descriptors
RDF145m	Radial Distribution Function - 145 / weighted by mass	RDF descriptors
RDF150m	Radial Distribution Function - 150 / weighted by mass	RDF descriptors
RDF155m	Radial Distribution Function - 155 / weighted by mass	RDF descriptors
RDF010v	Radial Distribution Function - 010 / weighted by van der Waals volume	RDF descriptors
RDF015v	Radial Distribution Function - 015 / weighted by van der Waals volume	RDF descriptors
RDF020v	Radial Distribution Function - 020 / weighted by van der Waals volume	RDF descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
RDF025v	Radial Distribution Function - 025 / weighted by van der Waals volume	RDF descriptors
RDF030v	Radial Distribution Function - 030 / weighted by van der Waals volume	RDF descriptors
RDF035v	Radial distribution function - 035 / weighted by van der waals volume	RDF descriptors
RDF040v	Radial distribution function - 040 / weighted by van der waals volume	RDF descriptors
RDF045v	Radial distribution function - 045 / weighted by van der waals volume	RDF descriptors
RDF050v	Radial distribution function - 050 / weighted by van der waals volume	RDF descriptors
RDF055v	Radial distribution function - 055 / weighted by van der waals volume	RDF descriptors
RDF060v	Radial distribution function - 060 / weighted by van der waals volume	RDF descriptors
RDF065v	Radial distribution function - 065 / weighted by van der waals volume	RDF descriptors
RDF070v	Radial distribution function - 070 / weighted by van der waals volume	RDF descriptors
RDF075v	Radial distribution function - 075 / weighted by van der waals volume	RDF descriptors
RDF080v	Radial distribution function - 080 / weighted by van der waals volume	RDF descriptors
RDF085v	Radial distribution function - 085 / weighted by van der waals volume	RDF descriptors
RDF090v	Radial distribution function - 090 / weighted by van der waals volume	RDF descriptors
RDF095v	Radial distribution function - 095 / weighted by van der waals volume	RDF descriptors
RDF100v	Radial distribution function - 100 / weighted by van der waals volume	RDF descriptors
RDF105v	Radial distribution function - 105 / weighted by van der waals volume	RDF descriptors
RDF110v	Radial distribution function - 110 / weighted by van der waals volume	RDF descriptors
RDF115v	Radial distribution function - 115 / weighted by van der waals volume	RDF descriptors
RDF120v	Radial distribution function - 120 / weighted by van der waals volume	RDF descriptors
RDF125v	Radial distribution function - 125 / weighted by van der waals volume	RDF descriptors
RDF130v	Radial distribution function - 130 / weighted by van der waals volume	RDF descriptors
RDF135v	Radial distribution function - 135 / weighted by van der waals volume	RDF descriptors
RDF140v	Radial distribution function - 140 / weighted by van der waals volume	RDF descriptors
RDF145v	Radial distribution function - 145 / weighted by van der waals volume	RDF descriptors
RDF150v	Radial distribution function - 150 / weighted by van der waals volume	RDF descriptors
RDF155v	Radial distribution function - 155 / weighted by van der waals volume	RDF descriptors
RDF010e	Radial distribution function - 010 / weighted by sanderson electronegativity	RDF descriptors
RDF015e	Radial distribution function - 015 / weighted by sanderson electronegativity	RDF descriptors
RDF020e	Radial distribution function - 020 / weighted by sanderson electronegativity	RDF descriptors
RDF025e	Radial distribution function - 025 / weighted by sanderson electronegativity	RDF descriptors
RDF030e	Radial distribution function - 030 / weighted by sanderson electronegativity	RDF descriptors
RDF035e	Radial distribution function - 035 / weighted by sanderson electronegativity	RDF descriptors
RDF040e	Radial distribution function - 040 / weighted by sanderson electronegativity	RDF descriptors
RDF045e	Radial distribution function - 045 / weighted by sanderson electronegativity	RDF descriptors
RDF050e	Radial distribution function - 050 / weighted by sanderson electronegativity	RDF descriptors
RDF055e	Radial distribution function - 055 / weighted by sanderson electronegativity	RDF descriptors
RDF060e	Radial distribution function - 060 / weighted by sanderson electronegativity	RDF descriptors
RDF065e	Radial distribution function - 065 / weighted by sanderson electronegativity	RDF descriptors
RDF070e	Radial distribution function - 070 / weighted by sanderson electronegativity	RDF descriptors
RDF075e	Radial distribution function - 075 / weighted by sanderson electronegativity	RDF descriptors
RDF080e	Radial distribution function - 080 / weighted by sanderson electronegativity	RDF descriptors
RDF085e	Radial distribution function - 085 / weighted by sanderson electronegativity	RDF descriptors
RDF090e	Radial distribution function - 090 / weighted by sanderson electronegativity	RDF descriptors
RDF095e	Radial distribution function - 095 / weighted by sanderson electronegativity	RDF descriptors
RDF100e	Radial distribution function - 100 / weighted by sanderson electronegativity	RDF descriptors
RDF105e	Radial distribution function - 105 / weighted by sanderson electronegativity	RDF descriptors
RDF110e	Radial distribution function - 110 / weighted by sanderson electronegativity	RDF descriptors
RDF115e	Radial distribution function - 115 / weighted by sanderson electronegativity	RDF descriptors
RDF120e	Radial distribution function - 120 / weighted by sanderson electronegativity	RDF descriptors
RDF125e	Radial distribution function - 125 / weighted by sanderson electronegativity	RDF descriptors
RDF130e	Radial distribution function - 130 / weighted by sanderson electronegativity	RDF descriptors
RDF135e	Radial distribution function - 135 / weighted by sanderson electronegativity	RDF descriptors
RDF140e	Radial distribution function - 140 / weighted by sanderson electronegativity	RDF descriptors
RDF145e	Radial distribution function - 145 / weighted by sanderson electronegativity	RDF descriptors
RDF150e	Radial distribution function - 150 / weighted by sanderson electronegativity	RDF descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
RDF155e	Radial distribution function - 155 / weighted by sanderson electronegativity	RDF descriptors
RDF010p	Radial distribution function - 010 / weighted by polarizability	RDF descriptors
RDF015p	Radial Distribution Function - 015 / weighted by polarizability	RDF descriptors
RDF020p	Radial Distribution Function - 020 / weighted by polarizability	RDF descriptors
RDF025p	Radial Distribution Function - 025 / weighted by polarizability	RDF descriptors
RDF030p	Radial Distribution Function - 030 / weighted by polarizability	RDF descriptors
RDF035p	Radial Distribution Function - 035 / weighted by polarizability	RDF descriptors
RDF040p	Radial Distribution Function - 040 / weighted by polarizability	RDF descriptors
RDF045p	Radial Distribution Function - 045 / weighted by polarizability	RDF descriptors
RDF050p	Radial Distribution Function - 050 / weighted by polarizability	RDF descriptors
RDF055p	Radial Distribution Function - 055 / weighted by polarizability	RDF descriptors
RDF060p	Radial Distribution Function - 060 / weighted by polarizability	RDF descriptors
RDF065p	Radial Distribution Function - 065 / weighted by polarizability	RDF descriptors
RDF070p	Radial Distribution Function - 070 / weighted by polarizability	RDF descriptors
RDF075p	Radial Distribution Function - 075 / weighted by polarizability	RDF descriptors
RDF080p	Radial Distribution Function - 080 / weighted by polarizability	RDF descriptors
RDF085p	Radial Distribution Function - 085 / weighted by polarizability	RDF descriptors
RDF090p	Radial Distribution Function - 090 / weighted by polarizability	RDF descriptors
RDF095p	Radial Distribution Function - 095 / weighted by polarizability	RDF descriptors
RDF100p	Radial Distribution Function - 100 / weighted by polarizability	RDF descriptors
RDF105p	Radial Distribution Function - 105 / weighted by polarizability	RDF descriptors
RDF110p	Radial Distribution Function - 110 / weighted by polarizability	RDF descriptors
RDF115p	Radial Distribution Function - 115 / weighted by polarizability	RDF descriptors
RDF120p	Radial Distribution Function - 120 / weighted by polarizability	RDF descriptors
RDF125p	Radial Distribution Function - 125 / weighted by polarizability	RDF descriptors
RDF130p	Radial Distribution Function - 130 / weighted by polarizability	RDF descriptors
RDF135p	Radial Distribution Function - 135 / weighted by polarizability	RDF descriptors
RDF140p	Radial Distribution Function - 140 / weighted by polarizability	RDF descriptors
RDF145p	Radial Distribution Function - 145 / weighted by polarizability	RDF descriptors
RDF150p	Radial Distribution Function - 150 / weighted by polarizability	RDF descriptors
RDF155p	Radial Distribution Function - 155 / weighted by polarizability	RDF descriptors
RDF010i	Radial Distribution Function - 010 / weighted by ionization potential	RDF descriptors
RDF015i	Radial Distribution Function - 015 / weighted by ionization potential	RDF descriptors
RDF020i	Radial Distribution Function - 020 / weighted by ionization potential	RDF descriptors
RDF025i	Radial Distribution Function - 025 / weighted by ionization potential	RDF descriptors
RDF030i	Radial Distribution Function - 030 / weighted by ionization potential	RDF descriptors
RDF035i	Radial Distribution Function - 035 / weighted by ionization potential	RDF descriptors
RDF040i	Radial Distribution Function - 040 / weighted by ionization potential	RDF descriptors
RDF045i	Radial Distribution Function - 045 / weighted by ionization potential	RDF descriptors
RDF050i	Radial Distribution Function - 050 / weighted by ionization potential	RDF descriptors
RDF055i	Radial Distribution Function - 055 / weighted by ionization potential	RDF descriptors
RDF060i	Radial Distribution Function - 060 / weighted by ionization potential	RDF descriptors
RDF065i	Radial Distribution Function - 065 / weighted by ionization potential	RDF descriptors
RDF070i	Radial Distribution Function - 070 / weighted by ionization potential	RDF descriptors
RDF075i	Radial Distribution Function - 075 / weighted by ionization potential	RDF descriptors
RDF080i	Radial Distribution Function - 080 / weighted by ionization potential	RDF descriptors
RDF085i	Radial Distribution Function - 085 / weighted by ionization potential	RDF descriptors
RDF090i	Radial Distribution Function - 090 / weighted by ionization potential	RDF descriptors
RDF095i	Radial Distribution Function - 095 / weighted by ionization potential	RDF descriptors
RDF100i	Radial Distribution Function - 100 / weighted by ionization potential	RDF descriptors
RDF105i	Radial Distribution Function - 105 / weighted by ionization potential	RDF descriptors
RDF110i	Radial Distribution Function - 110 / weighted by ionization potential	RDF descriptors
RDF115i	Radial Distribution Function - 115 / weighted by ionization potential	RDF descriptors
RDF120i	Radial Distribution Function - 120 / weighted by ionization potential	RDF descriptors
RDF125i	Radial Distribution Function - 125 / weighted by ionization potential	RDF descriptors
RDF130i	Radial Distribution Function - 130 / weighted by ionization potential	RDF descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
RDF135i	Radial Distribution Function - 135 / weighted by ionization potential	RDF descriptors
RDF140i	Radial Distribution Function - 140 / weighted by ionization potential	RDF descriptors
RDF145i	Radial Distribution Function - 145 / weighted by ionization potential	RDF descriptors
RDF150i	Radial Distribution Function - 150 / weighted by ionization potential	RDF descriptors
RDF155i	Radial Distribution Function - 155 / weighted by ionization potential	RDF descriptors
RDF010s	Radial Distribution Function - 010 / weighted by I-state	RDF descriptors
RDF015s	Radial Distribution Function - 015 / weighted by I-state	RDF descriptors
RDF020s	Radial Distribution Function - 020 / weighted by I-state	RDF descriptors
RDF025s	Radial Distribution Function - 025 / weighted by I-state	RDF descriptors
RDF030s	Radial Distribution Function - 030 / weighted by I-state	RDF descriptors
RDF035s	Radial Distribution Function - 035 / weighted by I-state	RDF descriptors
RDF040s	Radial Distribution Function - 040 / weighted by I-state	RDF descriptors
RDF045s	Radial Distribution Function - 045 / weighted by I-state	RDF descriptors
RDF050s	Radial Distribution Function - 050 / weighted by I-state	RDF descriptors
RDF055s	Radial Distribution Function - 055 / weighted by I-state	RDF descriptors
RDF060s	Radial Distribution Function - 060 / weighted by I-state	RDF descriptors
RDF065s	Radial Distribution Function - 065 / weighted by I-state	RDF descriptors
RDF070s	Radial Distribution Function - 070 / weighted by I-state	RDF descriptors
RDF075s	Radial Distribution Function - 075 / weighted by I-state	RDF descriptors
RDF080s	Radial Distribution Function - 080 / weighted by I-state	RDF descriptors
RDF085s	Radial Distribution Function - 085 / weighted by I-state	RDF descriptors
RDF090s	Radial Distribution Function - 090 / weighted by I-state	RDF descriptors
RDF095s	Radial Distribution Function - 095 / weighted by I-state	RDF descriptors
RDF100s	Radial Distribution Function - 100 / weighted by I-state	RDF descriptors
RDF105s	Radial Distribution Function - 105 / weighted by I-state	RDF descriptors
RDF110s	Radial Distribution Function - 110 / weighted by I-state	RDF descriptors
RDF115s	Radial Distribution Function - 115 / weighted by I-state	RDF descriptors
RDF120s	Radial Distribution Function - 120 / weighted by I-state	RDF descriptors
RDF125s	Radial Distribution Function - 125 / weighted by I-state	RDF descriptors
RDF130s	Radial Distribution Function - 130 / weighted by I-state	RDF descriptors
RDF135s	Radial Distribution Function - 135 / weighted by I-state	RDF descriptors
RDF140s	Radial Distribution Function - 140 / weighted by I-state	RDF descriptors
RDF145s	Radial Distribution Function - 145 / weighted by I-state	RDF descriptors
RDF150s	Radial Distribution Function - 150 / weighted by I-state	RDF descriptors
RDF155s	Radial Distribution Function - 155 / weighted by I-state	RDF descriptors
Mor02u	Signal 02 / unweighted	3D-MoRSE descriptors
Mor03u	Signal 03 / unweighted	3D-MoRSE descriptors
Mor04u	Signal 04 / unweighted	3D-MoRSE descriptors
Mor05u	Signal 05 / unweighted	3D-MoRSE descriptors
Mor06u	Signal 06 / unweighted	3D-MoRSE descriptors
Mor07u	Signal 07 / unweighted	3D-MoRSE descriptors
Mor08u	Signal 08 / unweighted	3D-MoRSE descriptors
Mor09u	Signal 09 / unweighted	3D-MoRSE descriptors
Mor10u	Signal 10 / unweighted	3D-MoRSE descriptors
Mor11u	Signal 11 / unweighted	3D-MoRSE descriptors
Mor12u	Signal 12 / unweighted	3D-MoRSE descriptors
Mor13u	Signal 13 / unweighted	3D-MoRSE descriptors
Mor14u	Signal 14 / unweighted	3D-MoRSE descriptors
Mor15u	Signal 15 / unweighted	3D-MoRSE descriptors
Mor16u	Signal 16 / unweighted	3D-MoRSE descriptors
Mor17u	Signal 17 / unweighted	3D-MoRSE descriptors
Mor18u	Signal 18 / unweighted	3D-MoRSE descriptors
Mor19u	Signal 19 / unweighted	3D-MoRSE descriptors
Mor20u	Signal 20 / unweighted	3D-MoRSE descriptors
Mor21u	Signal 21 / unweighted	3D-MoRSE descriptors
Mor22u	Signal 22 / unweighted	3D-MoRSE descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
Mor23u	Signal 23 / unweighted	3D-MoRSE descriptors
Mor24u	Signal 24 / unweighted	3D-MoRSE descriptors
Mor25u	Signal 25 / unweighted	3D-MoRSE descriptors
Mor26u	Signal 26 / unweighted	3D-MoRSE descriptors
Mor27u	Signal 27 / unweighted	3D-MoRSE descriptors
Mor28u	Signal 28 / unweighted	3D-MoRSE descriptors
Mor29u	Signal 29 / unweighted	3D-MoRSE descriptors
Mor30u	Signal 30 / unweighted	3D-MoRSE descriptors
Mor31u	Signal 31 / unweighted	3D-MoRSE descriptors
Mor32u	Signal 32 / unweighted	3D-MoRSE descriptors
Mor02m	Signal 02 / weighted by mass	3D-MoRSE descriptors
Mor03m	Signal 03 / weighted by mass	3D-MoRSE descriptors
Mor04m	Signal 04 / weighted by mass	3D-MoRSE descriptors
Mor05m	Signal 05 / weighted by mass	3D-MoRSE descriptors
Mor06m	Signal 06 / weighted by mass	3D-MoRSE descriptors
Mor07m	Signal 07 / weighted by mass	3D-MoRSE descriptors
Mor08m	Signal 08 / weighted by mass	3D-MoRSE descriptors
Mor09m	Signal 09 / weighted by mass	3D-MoRSE descriptors
Mor10m	Signal 10 / weighted by mass	3D-MoRSE descriptors
Mor11m	Signal 11 / weighted by mass	3D-MoRSE descriptors
Mor12m	Signal 12 / weighted by mass	3D-MoRSE descriptors
Mor13m	Signal 13 / weighted by mass	3D-MoRSE descriptors
Mor14m	Signal 14 / weighted by mass	3D-MoRSE descriptors
Mor15m	Signal 15 / weighted by mass	3D-MoRSE descriptors
Mor16m	Signal 16 / weighted by mass	3D-MoRSE descriptors
Mor17m	Signal 17 / weighted by mass	3D-MoRSE descriptors
Mor18m	Signal 18 / weighted by mass	3D-MoRSE descriptors
Mor19m	Signal 19 / weighted by mass	3D-MoRSE descriptors
Mor20m	Signal 20 / weighted by mass	3D-MoRSE descriptors
Mor21m	Signal 21 / weighted by mass	3D-MoRSE descriptors
Mor22m	Signal 22 / weighted by mass	3D-MoRSE descriptors
Mor23m	Signal 23 / weighted by mass	3D-MoRSE descriptors
Mor24m	Signal 24 / weighted by mass	3D-MoRSE descriptors
Mor25m	Signal 25 / weighted by mass	3D-MoRSE descriptors
Mor26m	Signal 26 / weighted by mass	3D-MoRSE descriptors
Mor27m	Signal 27 / weighted by mass	3D-MoRSE descriptors
Mor28m	Signal 28 / weighted by mass	3D-MoRSE descriptors
Mor29m	Signal 29 / weighted by mass	3D-MoRSE descriptors
Mor30m	Signal 30 / weighted by mass	3D-MoRSE descriptors
Mor31m	Signal 31 / weighted by mass	3D-MoRSE descriptors
Mor32m	Signal 32 / weighted by mass	3D-MoRSE descriptors
Mor02v	Signal 02 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor03v	Signal 03 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor04v	Signal 04 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor05v	Signal 05 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor06v	Signal 06 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor07v	Signal 07 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor08v	Signal 08 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor09v	Signal 09 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor10v	Signal 10 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor11v	Signal 11 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor12v	Signal 12 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor13v	Signal 13 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor14v	Signal 14 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor15v	Signal 15 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor16v	Signal 16 / weighted by van der Waals volume	3D-MoRSE descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
Mor17v	Signal 17 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor18v	Signal 18 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor19v	Signal 19 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor20v	Signal 20 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor21v	Signal 21 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor22v	Signal 22 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor23v	Signal 23 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor24v	Signal 24 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor25v	Signal 25 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor26v	Signal 26 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor27v	Signal 27 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor28v	Signal 28 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor29v	Signal 29 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor30v	Signal 30 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor31v	Signal 31 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor32v	Signal 32 / weighted by van der Waals volume	3D-MoRSE descriptors
Mor02e	Signal 02 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor03e	Signal 03 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor04e	Signal 04 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor05e	Signal 05 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor06e	Signal 06 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor07e	Signal 07 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor08e	Signal 08 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor09e	Signal 09 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor10e	Signal 10 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor11e	Signal 11 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor12e	Signal 12 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor13e	Signal 13 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor14e	Signal 14 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor15e	Signal 15 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor16e	Signal 16 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor17e	Signal 17 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor18e	Signal 18 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor19e	Signal 19 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor20e	Signal 20 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor21e	Signal 21 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor22e	Signal 22 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor23e	Signal 23 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor24e	Signal 24 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor25e	Signal 25 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor26e	Signal 26 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor27e	Signal 27 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor28e	Signal 28 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor29e	Signal 29 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor30e	Signal 30 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor31e	Signal 31 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor32e	Signal 32 / weighted by Sanderson electronegativity	3D-MoRSE descriptors
Mor02p	Signal 02 / weighted by polarizability	3D-MoRSE descriptors
Mor03p	Signal 03 / weighted by polarizability	3D-MoRSE descriptors
Mor04p	Signal 04 / weighted by polarizability	3D-MoRSE descriptors
Mor05p	Signal 05 / weighted by polarizability	3D-MoRSE descriptors
Mor06p	Signal 06 / weighted by polarizability	3D-MoRSE descriptors
Mor07p	Signal 07 / weighted by polarizability	3D-MoRSE descriptors
Mor08p	Signal 08 / weighted by polarizability	3D-MoRSE descriptors
Mor09p	Signal 09 / weighted by polarizability	3D-MoRSE descriptors
Mor10p	Signal 10 / weighted by polarizability	3D-MoRSE descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
Mor11p	Signal 11 / weighted by polarizability	3D-MoRSE descriptors
Mor12p	Signal 12 / weighted by polarizability	3D-MoRSE descriptors
Mor13p	Signal 13 / weighted by polarizability	3D-MoRSE descriptors
Mor14p	Signal 14 / weighted by polarizability	3D-MoRSE descriptors
Mor15p	Signal 15 / weighted by polarizability	3D-MoRSE descriptors
Mor16p	Signal 16 / weighted by polarizability	3D-MoRSE descriptors
Mor17p	Signal 17 / weighted by polarizability	3D-MoRSE descriptors
Mor18p	Signal 18 / weighted by polarizability	3D-MoRSE descriptors
Mor19p	Signal 19 / weighted by polarizability	3D-MoRSE descriptors
Mor20p	Signal 20 / weighted by polarizability	3D-MoRSE descriptors
Mor21p	Signal 21 / weighted by polarizability	3D-MoRSE descriptors
Mor22p	Signal 22 / weighted by polarizability	3D-MoRSE descriptors
Mor23p	Signal 23 / weighted by polarizability	3D-MoRSE descriptors
Mor24p	Signal 24 / weighted by polarizability	3D-MoRSE descriptors
Mor25p	Signal 25 / weighted by polarizability	3D-MoRSE descriptors
Mor26p	Signal 26 / weighted by polarizability	3D-MoRSE descriptors
Mor27p	Signal 27 / weighted by polarizability	3D-MoRSE descriptors
Mor28p	Signal 28 / weighted by polarizability	3D-MoRSE descriptors
Mor29p	Signal 29 / weighted by polarizability	3D-MoRSE descriptors
Mor30p	Signal 30 / weighted by polarizability	3D-MoRSE descriptors
Mor31p	Signal 31 / weighted by polarizability	3D-MoRSE descriptors
Mor32p	Signal 32 / weighted by polarizability	3D-MoRSE descriptors
Mor02i	Signal 02 / weighted by ionization potential	3D-MoRSE descriptors
Mor03i	Signal 03 / weighted by ionization potential	3D-MoRSE descriptors
Mor04i	Signal 04 / weighted by ionization potential	3D-MoRSE descriptors
Mor05i	Signal 05 / weighted by ionization potential	3D-MoRSE descriptors
Mor06i	Signal 06 / weighted by ionization potential	3D-MoRSE descriptors
Mor07i	Signal 07 / weighted by ionization potential	3D-MoRSE descriptors
Mor08i	Signal 08 / weighted by ionization potential	3D-MoRSE descriptors
Mor09i	Signal 09 / weighted by ionization potential	3D-MoRSE descriptors
Mor10i	Signal 10 / weighted by ionization potential	3D-MoRSE descriptors
Mor11i	Signal 11 / weighted by ionization potential	3D-MoRSE descriptors
Mor12i	Signal 12 / weighted by ionization potential	3D-MoRSE descriptors
Mor13i	Signal 13 / weighted by ionization potential	3D-MoRSE descriptors
Mor14i	Signal 14 / weighted by ionization potential	3D-MoRSE descriptors
Mor15i	Signal 15 / weighted by ionization potential	3D-MoRSE descriptors
Mor16i	Signal 16 / weighted by ionization potential	3D-MoRSE descriptors
Mor17i	Signal 17 / weighted by ionization potential	3D-MoRSE descriptors
Mor18i	Signal 18 / weighted by ionization potential	3D-MoRSE descriptors
Mor19i	Signal 19 / weighted by ionization potential	3D-MoRSE descriptors
Mor20i	Signal 20 / weighted by ionization potential	3D-MoRSE descriptors
Mor21i	Signal 21 / weighted by ionization potential	3D-MoRSE descriptors
Mor22i	Signal 22 / weighted by ionization potential	3D-MoRSE descriptors
Mor23i	Signal 23 / weighted by ionization potential	3D-MoRSE descriptors
Mor24i	Signal 24 / weighted by ionization potential	3D-MoRSE descriptors
Mor25i	Signal 25 / weighted by ionization potential	3D-MoRSE descriptors
Mor26i	Signal 26 / weighted by ionization potential	3D-MoRSE descriptors
Mor27i	Signal 27 / weighted by ionization potential	3D-MoRSE descriptors
Mor28i	Signal 28 / weighted by ionization potential	3D-MoRSE descriptors
Mor29i	Signal 29 / weighted by ionization potential	3D-MoRSE descriptors
Mor30i	Signal 30 / weighted by ionization potential	3D-MoRSE descriptors
Mor31i	Signal 31 / weighted by ionization potential	3D-MoRSE descriptors
Mor32i	Signal 32 / weighted by ionization potential	3D-MoRSE descriptors
Mor02s	Signal 02 / weighted by I-state	3D-MoRSE descriptors
Mor03s	Signal 03 / weighted by I-state	3D-MoRSE descriptors
Mor04s	Signal 04 / weighted by I-state	3D-MoRSE descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
Mor05s	Signal 05 / weighted by I-state	3D-MoRSE descriptors
Mor06s	Signal 06 / weighted by I-state	3D-MoRSE descriptors
Mor07s	Signal 07 / weighted by I-state	3D-MoRSE descriptors
Mor08s	Signal 08 / weighted by I-state	3D-MoRSE descriptors
Mor09s	Signal 09 / weighted by I-state	3D-MoRSE descriptors
Mor10s	Signal 10 / weighted by I-state	3D-MoRSE descriptors
Mor11s	Signal 11 / weighted by I-state	3D-MoRSE descriptors
Mor12s	Signal 12 / weighted by I-state	3D-MoRSE descriptors
Mor13s	Signal 13 / weighted by I-state	3D-MoRSE descriptors
Mor14s	Signal 14 / weighted by I-state	3D-MoRSE descriptors
Mor15s	Signal 15 / weighted by I-state	3D-MoRSE descriptors
Mor16s	Signal 16 / weighted by I-state	3D-MoRSE descriptors
Mor17s	Signal 17 / weighted by I-state	3D-MoRSE descriptors
Mor18s	Signal 18 / weighted by I-state	3D-MoRSE descriptors
Mor19s	Signal 19 / weighted by I-state	3D-MoRSE descriptors
Mor20s	Signal 20 / weighted by I-state	3D-MoRSE descriptors
Mor21s	Signal 21 / weighted by I-state	3D-MoRSE descriptors
Mor22s	Signal 22 / weighted by I-state	3D-MoRSE descriptors
Mor23s	Signal 23 / weighted by I-state	3D-MoRSE descriptors
Mor24s	Signal 24 / weighted by I-state	3D-MoRSE descriptors
Mor25s	Signal 25 / weighted by I-state	3D-MoRSE descriptors
Mor26s	Signal 26 / weighted by I-state	3D-MoRSE descriptors
Mor27s	Signal 27 / weighted by I-state	3D-MoRSE descriptors
Mor28s	Signal 28 / weighted by I-state	3D-MoRSE descriptors
Mor29s	Signal 29 / weighted by I-state	3D-MoRSE descriptors
Mor30s	Signal 30 / weighted by I-state	3D-MoRSE descriptors
Mor31s	Signal 31 / weighted by I-state	3D-MoRSE descriptors
Mor32s	Signal 32 / weighted by I-state	3D-MoRSE descriptors
L1u	1st component size directional WHIM index / unweighted	WHIM descriptors
L2u	2nd component size directional WHIM index / unweighted	WHIM descriptors
L3u	3rd component size directional WHIM index / unweighted	WHIM descriptors
P1u	1st component shape directional WHIM index / unweighted	WHIM descriptors
P2u	2nd component shape directional WHIM index / unweighted	WHIM descriptors
G1u	1st component symmetry directional WHIM index / unweighted	WHIM descriptors
G2u	2nd component symmetry directional WHIM index / unweighted	WHIM descriptors
G3u	3rd component symmetry directional WHIM index / unweighted	WHIM descriptors
E1u	1st component accessibility directional WHIM index / unweighted	WHIM descriptors
E2u	2nd component accessibility directional WHIM index / unweighted	WHIM descriptors
E3u	3rd component accessibility directional WHIM index / unweighted	WHIM descriptors
L1m	1st component size directional WHIM index / weighted by mass	WHIM descriptors
L2m	2nd component size directional WHIM index / weighted by mass	WHIM descriptors
L3m	3rd component size directional WHIM index / weighted by mass	WHIM descriptors
P1m	1st component shape directional WHIM index / weighted by mass	WHIM descriptors
P2m	2nd component shape directional WHIM index / weighted by mass	WHIM descriptors
G1m	1st component symmetry directional WHIM index / weighted by mass	WHIM descriptors
G2m	2nd component symmetry directional WHIM index / weighted by mass	WHIM descriptors
G3m	3rd component symmetry directional WHIM index / weighted by mass	WHIM descriptors
E1m	1st component accessibility directional WHIM index / weighted by mass	WHIM descriptors
E2m	2nd component accessibility directional WHIM index / weighted by mass	WHIM descriptors
E3m	3rd component accessibility directional WHIM index / weighted by mass	WHIM descriptors
L1v	1st component size directional WHIM index / weighted by van der Waals volume	WHIM descriptors
L2v	2nd component size directional WHIM index / weighted by van der Waals volume	WHIM descriptors
L3v	3rd component size directional WHIM index / weighted by van der Waals volume	WHIM descriptors
P1v	1st component shape directional WHIM index / weighted by van der Waals volume	WHIM descriptors
P2v	2nd component shape directional WHIM index / weighted by van der Waals volume	WHIM descriptors
G1v	1st component symmetry directional WHIM index / weighted by van der Waals volume	WHIM descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
G2v	2nd component symmetry directional WHIM index / weighted by van der Waals volume	WHIM descriptors
G3v	3rd component symmetry directional WHIM index / weighted by van der Waals volume	WHIM descriptors
E1v	1st component accessibility directional WHIM index / weighted by van der Waals volume	WHIM descriptors
E2v	2nd component accessibility directional WHIM index / weighted by van der Waals volume	WHIM descriptors
E3v	3rd component accessibility directional WHIM index / weighted by van der Waals volume	WHIM descriptors
L1e	1st component size directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
L2e	2nd component size directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
L3e	3rd component size directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
P1e	1st component shape directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
P2e	2nd component shape directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
G1e	1st component symmetry directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
G2e	2nd component symmetry directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
G3e	3rd component symmetry directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
E1e	1st component accessibility directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
E2e	2nd component accessibility directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
E3e	3rd component accessibility directional WHIM index / weighted by Sanderson electronegativity	WHIM descriptors
L1p	1st component size directional WHIM index / weighted by polarizability	WHIM descriptors
L2p	2nd component size directional WHIM index / weighted by polarizability	WHIM descriptors
L3p	3rd component size directional WHIM index / weighted by polarizability	WHIM descriptors
P1p	1st component shape directional WHIM index / weighted by polarizability	WHIM descriptors
P2p	2nd component shape directional WHIM index / weighted by polarizability	WHIM descriptors
G1p	1st component symmetry directional WHIM index / weighted by polarizability	WHIM descriptors
G2p	2nd component symmetry directional WHIM index / weighted by polarizability	WHIM descriptors
G3p	3rd component symmetry directional WHIM index / weighted by polarizability	WHIM descriptors
E1p	1st component accessibility directional WHIM index / weighted by polarizability	WHIM descriptors
E2p	2nd component accessibility directional WHIM index / weighted by polarizability	WHIM descriptors
E3p	3rd component accessibility directional WHIM index / weighted by polarizability	WHIM descriptors
L1i	1st component size directional WHIM index / weighted by ionization potential	WHIM descriptors
L2i	2nd component size directional WHIM index / weighted by ionization potential	WHIM descriptors
L3i	3rd component size directional WHIM index / weighted by ionization potential	WHIM descriptors
P1i	1st component shape directional WHIM index / weighted by ionization potential	WHIM descriptors
P2i	2nd component shape directional WHIM index / weighted by ionization potential	WHIM descriptors
G1i	1st component symmetry directional WHIM index / weighted by ionization potential	WHIM descriptors
G2i	2nd component symmetry directional WHIM index / weighted by ionization potential	WHIM descriptors
G3i	3rd component symmetry directional WHIM index / weighted by ionization potential	WHIM descriptors
E1i	1st component accessibility directional WHIM index / weighted by ionization potential	WHIM descriptors
E2i	2nd component accessibility directional WHIM index / weighted by ionization potential	WHIM descriptors
E3i	3rd component accessibility directional WHIM index / weighted by ionization potential	WHIM descriptors
L1s	1st component size directional WHIM index / weighted by I-state	WHIM descriptors
L2s	2nd component size directional WHIM index / weighted by I-state	WHIM descriptors
L3s	3rd component size directional WHIM index / weighted by I-state	WHIM descriptors
P1s	1st component shape directional WHIM index / weighted by I-state	WHIM descriptors
P2s	2nd component shape directional WHIM index / weighted by I-state	WHIM descriptors
G1s	1st component symmetry directional WHIM index / weighted by I-state	WHIM descriptors
G2s	2nd component symmetry directional WHIM index / weighted by I-state	WHIM descriptors
G3s	3rd component symmetry directional WHIM index / weighted by I-state	WHIM descriptors
E1s	1st component accessibility directional WHIM index / weighted by I-state	WHIM descriptors
E2s	2nd component accessibility directional WHIM index / weighted by I-state	WHIM descriptors
E3s	3rd component accessibility directional WHIM index / weighted by I-state	WHIM descriptors
Tu	T total size index / unweighted	WHIM descriptors
Tm	T total size index / weighted by mass	WHIM descriptors
Tv	T total size index / weighted by van der Waals volume	WHIM descriptors
Te	T total size index / weighted by Sanderson electronegativity	WHIM descriptors
Tp	T total size index / weighted by polarizability	WHIM descriptors
Ti	T total size index / weighted by ionization potential	WHIM descriptors
Ts	T total size index / weighted by I-state	WHIM descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
Au	A total size index / unweighted	WHIM descriptors
Am	A total size index / weighted by mass	WHIM descriptors
Av	A total size index / weighted by van der Waals volume	WHIM descriptors
Ae	A total size index / weighted by Sanderson electronegativity	WHIM descriptors
Ap	A total size index / weighted by polarizability	WHIM descriptors
Ai	A total size index / weighted by ionization potential	WHIM descriptors
As	A total size index / weighted by I-state	WHIM descriptors
Gu	Total symmetry index / unweighted	WHIM descriptors
Gm	Total symmetry index / weighted by mass	WHIM descriptors
Ku	K global shape index / unweighted	WHIM descriptors
Km	K global shape index / weighted by mass	WHIM descriptors
Kv	K global shape index / weighted by van der Waals volume	WHIM descriptors
Ke	K global shape index / weighted by Sanderson electronegativity	WHIM descriptors
Kp	K global shape index / weighted by polarizability	WHIM descriptors
Ki	K global shape index / weighted by ionization potential	WHIM descriptors
Ks	K global shape index / weighted by I-state	WHIM descriptors
Du	D total accessibility index / unweighted	WHIM descriptors
Dm	D total accessibility index / weighted by mass	WHIM descriptors
Dv	D total accessibility index / weighted by van der Waals volume	WHIM descriptors
De	D total accessibility index / weighted by Sanderson electronegativity	WHIM descriptors
Dp	D total accessibility index / weighted by polarizability	WHIM descriptors
Di	D total accessibility index / weighted by ionization potential	WHIM descriptors
Ds	D total accessibility index / weighted by I-state	WHIM descriptors
Vu	V total size index / unweighted	WHIM descriptors
Vm	V total size index / weighted by mass	WHIM descriptors
Vv	V total size index / weighted by van der Waals volume	WHIM descriptors
Ve	V total size index / weighted by Sanderson electronegativity	WHIM descriptors
Vp	V total size index / weighted by polarizability	WHIM descriptors
Vi	V total size index / weighted by ionization potential	WHIM descriptors
Vs	V total size index / weighted by I-state	WHIM descriptors
ITH	Total information content on the leverage equality	GETAWAY descriptors
ISH	Standardized information content on the leverage equality	GETAWAY descriptors
HIC	Mean information content on the leverage magnitude	GETAWAY descriptors
HGM	Geometric mean on the leverage magnitude	GETAWAY descriptors
H1u	H autocorrelation of lag 1 / unweighted	GETAWAY descriptors
H2u	H autocorrelation of lag 2 / unweighted	GETAWAY descriptors
H3u	H autocorrelation of lag 3 / unweighted	GETAWAY descriptors
H4u	H autocorrelation of lag 4 / unweighted	GETAWAY descriptors
H5u	H autocorrelation of lag 5 / unweighted	GETAWAY descriptors
H6u	H autocorrelation of lag 6 / unweighted	GETAWAY descriptors
H7u	H autocorrelation of lag 7 / unweighted	GETAWAY descriptors
H8u	H autocorrelation of lag 8 / unweighted	GETAWAY descriptors
HTu	H total index / unweighted	GETAWAY descriptors
HATS0u	Leverage-weighted autocorrelation of lag 0 / unweighted	GETAWAY descriptors
HATS1u	Leverage-weighted autocorrelation of lag 1 / unweighted	GETAWAY descriptors
HATS2u	Leverage-weighted autocorrelation of lag 2 / unweighted	GETAWAY descriptors
HATS3u	Leverage-weighted autocorrelation of lag 3 / unweighted	GETAWAY descriptors
HATS4u	Leverage-weighted autocorrelation of lag 4 / unweighted	GETAWAY descriptors
HATS5u	Leverage-weighted autocorrelation of lag 5 / unweighted	GETAWAY descriptors
HATS6u	Leverage-weighted autocorrelation of lag 6 / unweighted	GETAWAY descriptors
HATS7u	Leverage-weighted autocorrelation of lag 7 / unweighted	GETAWAY descriptors
HATS8u	Leverage-weighted autocorrelation of lag 8 / unweighted	GETAWAY descriptors
H0m	H autocorrelation of lag 0 / weighted by mass	GETAWAY descriptors
H1m	H autocorrelation of lag 1 / weighted by mass	GETAWAY descriptors
H2m	H autocorrelation of lag 2 / weighted by mass	GETAWAY descriptors
H3m	H autocorrelation of lag 3 / weighted by mass	GETAWAY descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
H4m	H autocorrelation of lag 4 / weighted by mass	GETAWAY descriptors
H5m	H autocorrelation of lag 5 / weighted by mass	GETAWAY descriptors
H6m	H autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors
H7m	H autocorrelation of lag 7 / weighted by mass	GETAWAY descriptors
H8m	H autocorrelation of lag 8 / weighted by mass	GETAWAY descriptors
HTm	H total index / weighted by mass	GETAWAY descriptors
HATS0m	Leverage-weighted autocorrelation of lag 0 / weighted by mass	GETAWAY descriptors
HATS1m	Leverage-weighted autocorrelation of lag 1 / weighted by mass	GETAWAY descriptors
HATS2m	Leverage-weighted autocorrelation of lag 2 / weighted by mass	GETAWAY descriptors
HATS3m	Leverage-weighted autocorrelation of lag 3 / weighted by mass	GETAWAY descriptors
HATS4m	Leverage-weighted autocorrelation of lag 4 / weighted by mass	GETAWAY descriptors
HATS5m	Leverage-weighted autocorrelation of lag 5 / weighted by mass	GETAWAY descriptors
HATS6m	Leverage-weighted autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors
HATS7m	Leverage-weighted autocorrelation of lag 7 / weighted by mass	GETAWAY descriptors
HATS8m	Leverage-weighted autocorrelation of lag 8 / weighted by mass	GETAWAY descriptors
HATSm	Leverage-weighted total index / weighted by mass	GETAWAY descriptors
H0v	H autocorrelation of lag 0 / weighted by van der Waals volume	GETAWAY descriptors
H1v	H autocorrelation of lag 1 / weighted by van der Waals volume	GETAWAY descriptors
H2v	H autocorrelation of lag 2 / weighted by van der Waals volume	GETAWAY descriptors
H3v	H autocorrelation of lag 3 / weighted by van der Waals volume	GETAWAY descriptors
H4v	H autocorrelation of lag 4 / weighted by van der Waals volume	GETAWAY descriptors
H5v	H autocorrelation of lag 5 / weighted by van der Waals volume	GETAWAY descriptors
H6v	H autocorrelation of lag 6 / weighted by van der Waals volume	GETAWAY descriptors
H7v	H autocorrelation of lag 7 / weighted by van der Waals volume	GETAWAY descriptors
H8v	H autocorrelation of lag 8 / weighted by van der Waals volume	GETAWAY descriptors
HTv	H total index / weighted by van der Waals volume	GETAWAY descriptors
HATS0v	Leverage-weighted autocorrelation of lag 0 / weighted by van der Waals volume	GETAWAY descriptors
HATS1v	Leverage-weighted autocorrelation of lag 1 / weighted by van der Waals volume	GETAWAY descriptors
HATS2v	Leverage-weighted autocorrelation of lag 2 / weighted by van der Waals volume	GETAWAY descriptors
HATS3v	Leverage-weighted autocorrelation of lag 3 / weighted by van der Waals volume	GETAWAY descriptors
HATS4v	Leverage-weighted autocorrelation of lag 4 / weighted by van der Waals volume	GETAWAY descriptors
HATS5v	Leverage-weighted autocorrelation of lag 5 / weighted by van der Waals volume	GETAWAY descriptors
HATS6v	Leverage-weighted autocorrelation of lag 6 / weighted by van der Waals volume	GETAWAY descriptors
HATS7v	Leverage-weighted autocorrelation of lag 7 / weighted by van der Waals volume	GETAWAY descriptors
HATS8v	Leverage-weighted autocorrelation of lag 8 / weighted by van der Waals volume	GETAWAY descriptors
HATSv	Leverage-weighted total index / weighted by van der Waals volume	GETAWAY descriptors
H0e	H autocorrelation of lag 0 / weighted by Sanderson electronegativity	GETAWAY descriptors
H1e	H autocorrelation of lag 1 / weighted by Sanderson electronegativity	GETAWAY descriptors
H2e	H autocorrelation of lag 2 / weighted by Sanderson electronegativity	GETAWAY descriptors
H3e	H autocorrelation of lag 3 / weighted by Sanderson electronegativity	GETAWAY descriptors
H4e	H autocorrelation of lag 4 / weighted by Sanderson electronegativity	GETAWAY descriptors
H5e	H autocorrelation of lag 5 / weighted by Sanderson electronegativity	GETAWAY descriptors
H6e	H autocorrelation of lag 6 / weighted by Sanderson electronegativity	GETAWAY descriptors
H7e	H autocorrelation of lag 7 / weighted by Sanderson electronegativity	GETAWAY descriptors
H8e	H autocorrelation of lag 8 / weighted by Sanderson electronegativity	GETAWAY descriptors
HTe	H total index / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS0e	Leverage-weighted autocorrelation of lag 0 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS1e	Leverage-weighted autocorrelation of lag 1 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS2e	Leverage-weighted autocorrelation of lag 2 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS3e	Leverage-weighted autocorrelation of lag 3 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS4e	Leverage-weighted autocorrelation of lag 4 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS5e	Leverage-weighted autocorrelation of lag 5 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS6e	Leverage-weighted autocorrelation of lag 6 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS7e	Leverage-weighted autocorrelation of lag 7 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATS8e	Leverage-weighted autocorrelation of lag 8 / weighted by Sanderson electronegativity	GETAWAY descriptors
HATSe	Leverage-weighted total index / weighted by Sanderson electronegativity	GETAWAY descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
H0p	H autocorrelation of lag 0 / weighted by polarizability	GETAWAY descriptors
H1p	H autocorrelation of lag 1 / weighted by polarizability	GETAWAY descriptors
H2p	H autocorrelation of lag 2 / weighted by polarizability	GETAWAY descriptors
H3p	H autocorrelation of lag 3 / weighted by polarizability	GETAWAY descriptors
H4p	H autocorrelation of lag 4 / weighted by polarizability	GETAWAY descriptors
H5p	H autocorrelation of lag 5 / weighted by polarizability	GETAWAY descriptors
H6p	H autocorrelation of lag 6 / weighted by polarizability	GETAWAY descriptors
H7p	H autocorrelation of lag 7 / weighted by polarizability	GETAWAY descriptors
H8p	H autocorrelation of lag 8 / weighted by polarizability	GETAWAY descriptors
HTp	H total index / weighted by polarizability	GETAWAY descriptors
HATS0p	Leverage-weighted autocorrelation of lag 0 / weighted by polarizability	GETAWAY descriptors
HATS1p	Leverage-weighted autocorrelation of lag 1 / weighted by polarizability	GETAWAY descriptors
HATS2p	Leverage-weighted autocorrelation of lag 2 / weighted by polarizability	GETAWAY descriptors
HATS3p	Leverage-weighted autocorrelation of lag 3 / weighted by polarizability	GETAWAY descriptors
HATS4p	Leverage-weighted autocorrelation of lag 4 / weighted by polarizability	GETAWAY descriptors
HATS5p	Leverage-weighted autocorrelation of lag 5 / weighted by polarizability	GETAWAY descriptors
HATS6p	Leverage-weighted autocorrelation of lag 6 / weighted by polarizability	GETAWAY descriptors
HATS7p	Leverage-weighted autocorrelation of lag 7 / weighted by polarizability	GETAWAY descriptors
HATS8p	Leverage-weighted autocorrelation of lag 8 / weighted by polarizability	GETAWAY descriptors
HATSp	Leverage-weighted total index / weighted by polarizability	GETAWAY descriptors
H0i	H autocorrelation of lag 0 / weighted by ionization potential	GETAWAY descriptors
H1i	H autocorrelation of lag 1 / weighted by ionization potential	GETAWAY descriptors
H2i	H autocorrelation of lag 2 / weighted by ionization potential	GETAWAY descriptors
H3i	H autocorrelation of lag 3 / weighted by ionization potential	GETAWAY descriptors
H4i	H autocorrelation of lag 4 / weighted by ionization potential	GETAWAY descriptors
H5i	H autocorrelation of lag 5 / weighted by ionization potential	GETAWAY descriptors
H6i	H autocorrelation of lag 6 / weighted by ionization potential	GETAWAY descriptors
H7i	H autocorrelation of lag 7 / weighted by ionization potential	GETAWAY descriptors
H8i	H autocorrelation of lag 8 / weighted by ionization potential	GETAWAY descriptors
HTi	H total index / weighted by ionization potential	GETAWAY descriptors
HATS0i	Leverage-weighted autocorrelation of lag 0 / weighted by ionization potential	GETAWAY descriptors
HATS1i	Leverage-weighted autocorrelation of lag 1 / weighted by ionization potential	GETAWAY descriptors
HATS2i	Leverage-weighted autocorrelation of lag 2 / weighted by ionization potential	GETAWAY descriptors
HATS3i	Leverage-weighted autocorrelation of lag 3 / weighted by ionization potential	GETAWAY descriptors
HATS4i	Leverage-weighted autocorrelation of lag 4 / weighted by ionization potential	GETAWAY descriptors
HATS5i	Leverage-weighted autocorrelation of lag 5 / weighted by ionization potential	GETAWAY descriptors
HATS6i	Leverage-weighted autocorrelation of lag 6 / weighted by ionization potential	GETAWAY descriptors
HATS7i	Leverage-weighted autocorrelation of lag 7 / weighted by ionization potential	GETAWAY descriptors
HATS8i	Leverage-weighted autocorrelation of lag 8 / weighted by ionization potential	GETAWAY descriptors
HATSi	Leverage-weighted total index / weighted by ionization potential	GETAWAY descriptors
H0s	H autocorrelation of lag 0 / weighted by I-state	GETAWAY descriptors
H1s	H autocorrelation of lag 1 / weighted by I-state	GETAWAY descriptors
H2s	H autocorrelation of lag 2 / weighted by I-state	GETAWAY descriptors
H3s	H autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
H4s	H autocorrelation of lag 4 / weighted by I-state	GETAWAY descriptors
H5s	H autocorrelation of lag 5 / weighted by I-state	GETAWAY descriptors
H6s	H autocorrelation of lag 6 / weighted by I-state	GETAWAY descriptors
H7s	H autocorrelation of lag 7 / weighted by I-state	GETAWAY descriptors
H8s	H autocorrelation of lag 8 / weighted by I-state	GETAWAY descriptors
HTs	H total index / weighted by I-state	GETAWAY descriptors
HATS0s	Leverage-weighted autocorrelation of lag 0 / weighted by I-state	GETAWAY descriptors
HATS1s	Leverage-weighted autocorrelation of lag 1 / weighted by I-state	GETAWAY descriptors
HATS2s	Leverage-weighted autocorrelation of lag 2 / weighted by I-state	GETAWAY descriptors
HATS3s	Leverage-weighted autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
HATS4s	Leverage-weighted autocorrelation of lag 4 / weighted by I-state	GETAWAY descriptors
HATS5s	Leverage-weighted autocorrelation of lag 5 / weighted by I-state	GETAWAY descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
HATS6s	Leverage-weighted autocorrelation of lag 6 / weighted by I-state	GETAWAY descriptors
HATS7s	Leverage-weighted autocorrelation of lag 7 / weighted by I-state	GETAWAY descriptors
HATS8s	Leverage-weighted autocorrelation of lag 8 / weighted by I-state	GETAWAY descriptors
HATSs	Leverage-weighted total index / weighted by I-state	GETAWAY descriptors
RCON	Randic-type R matrix connectivity	GETAWAY descriptors
RARS	R matrix average row sum	GETAWAY descriptors
REIG	First eigenvalue of the R matrix	GETAWAY descriptors
R1u	R autocorrelation of lag 1 / unweighted	GETAWAY descriptors
R2u	R autocorrelation of lag 2 / unweighted	GETAWAY descriptors
R3u	R autocorrelation of lag 3 / unweighted	GETAWAY descriptors
R4u	R autocorrelation of lag 4 / unweighted	GETAWAY descriptors
R5u	R autocorrelation of lag 5 / unweighted	GETAWAY descriptors
R6u	R autocorrelation of lag 6 / unweighted	GETAWAY descriptors
R7u	R autocorrelation of lag 7 / unweighted	GETAWAY descriptors
R8u	R autocorrelation of lag 8 / unweighted	GETAWAY descriptors
RTu	R total index / unweighted	GETAWAY descriptors
R1u+	R maximal autocorrelation of lag 1 / unweighted	GETAWAY descriptors
R2u+	R maximal autocorrelation of lag 2 / unweighted	GETAWAY descriptors
R3u+	R maximal autocorrelation of lag 3 / unweighted	GETAWAY descriptors
R4u+	R maximal autocorrelation of lag 4 / unweighted	GETAWAY descriptors
R5u+	R maximal autocorrelation of lag 5 / unweighted	GETAWAY descriptors
R6u+	R maximal autocorrelation of lag 6 / unweighted	GETAWAY descriptors
R7u+	R maximal autocorrelation of lag 7 / unweighted	GETAWAY descriptors
R8u+	R maximal autocorrelation of lag 8 / unweighted	GETAWAY descriptors
RTu+	R maximal index / unweighted	GETAWAY descriptors
R1m	R autocorrelation of lag 1 / weighted by mass	GETAWAY descriptors
R2m	R autocorrelation of lag 2 / weighted by mass	GETAWAY descriptors
R3m	R autocorrelation of lag 3 / weighted by mass	GETAWAY descriptors
R4m	R autocorrelation of lag 4 / weighted by mass	GETAWAY descriptors
R5m	R autocorrelation of lag 5 / weighted by mass	GETAWAY descriptors
R6m	R autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors
R7m	R autocorrelation of lag 7 / weighted by mass	GETAWAY descriptors
R8m	R autocorrelation of lag 8 / weighted by mass	GETAWAY descriptors
RTm	R total index / weighted by mass	GETAWAY descriptors
R1m+	R maximal autocorrelation of lag 1 / weighted by mass	GETAWAY descriptors
R2m+	R maximal autocorrelation of lag 2 / weighted by mass	GETAWAY descriptors
R3m+	R maximal autocorrelation of lag 3 / weighted by mass	GETAWAY descriptors
R4m+	R maximal autocorrelation of lag 4 / weighted by mass	GETAWAY descriptors
R5m+	R maximal autocorrelation of lag 5 / weighted by mass	GETAWAY descriptors
R6m+	R maximal autocorrelation of lag 6 / weighted by mass	GETAWAY descriptors
R7m+	R maximal autocorrelation of lag 7 / weighted by mass	GETAWAY descriptors
R8m+	R maximal autocorrelation of lag 8 / weighted by mass	GETAWAY descriptors
RTm+	R maximal index / weighted by mass	GETAWAY descriptors
R1v	R autocorrelation of lag 1 / weighted by van der Waals volume	GETAWAY descriptors
R2v	R autocorrelation of lag 2 / weighted by van der Waals volume	GETAWAY descriptors
R3v	R autocorrelation of lag 3 / weighted by van der Waals volume	GETAWAY descriptors
R4v	R autocorrelation of lag 4 / weighted by van der Waals volume	GETAWAY descriptors
R5v	R autocorrelation of lag 5 / weighted by van der Waals volume	GETAWAY descriptors
R6v	R autocorrelation of lag 6 / weighted by van der Waals volume	GETAWAY descriptors
R7v	R autocorrelation of lag 7 / weighted by van der Waals volume	GETAWAY descriptors
R8v	R autocorrelation of lag 8 / weighted by van der Waals volume	GETAWAY descriptors
RTv	R total index / weighted by van der Waals volume	GETAWAY descriptors
R1v+	R maximal autocorrelation of lag 1 / weighted by van der Waals volume	GETAWAY descriptors
R2v+	R maximal autocorrelation of lag 2 / weighted by van der Waals volume	GETAWAY descriptors
R3v+	R maximal autocorrelation of lag 3 / weighted by van der Waals volume	GETAWAY descriptors
R4v+	R maximal autocorrelation of lag 4 / weighted by van der Waals volume	GETAWAY descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
R5v+	R maximal autocorrelation of lag 5 / weighted by van der Waals volume	GETAWAY descriptors
R6v+	R maximal autocorrelation of lag 6 / weighted by van der Waals volume	GETAWAY descriptors
R7v+	R maximal autocorrelation of lag 7 / weighted by van der Waals volume	GETAWAY descriptors
R8v+	R maximal autocorrelation of lag 8 / weighted by van der Waals volume	GETAWAY descriptors
RTv+	R maximal index / weighted by van der Waals volume	GETAWAY descriptors
R1e	R autocorrelation of lag 1 / weighted by Sanderson electronegativity	GETAWAY descriptors
R2e	R autocorrelation of lag 2 / weighted by Sanderson electronegativity	GETAWAY descriptors
R3e	R autocorrelation of lag 3 / weighted by Sanderson electronegativity	GETAWAY descriptors
R4e	R autocorrelation of lag 4 / weighted by Sanderson electronegativity	GETAWAY descriptors
R5e	R autocorrelation of lag 5 / weighted by Sanderson electronegativity	GETAWAY descriptors
R6e	R autocorrelation of lag 6 / weighted by Sanderson electronegativity	GETAWAY descriptors
R7e	R autocorrelation of lag 7 / weighted by Sanderson electronegativity	GETAWAY descriptors
R8e	R autocorrelation of lag 8 / weighted by Sanderson electronegativity	GETAWAY descriptors
RTe	R total index / weighted by Sanderson electronegativity	GETAWAY descriptors
R1e+	R maximal autocorrelation of lag 1 / weighted by Sanderson electronegativity	GETAWAY descriptors
R2e+	R maximal autocorrelation of lag 2 / weighted by Sanderson electronegativity	GETAWAY descriptors
R3e+	R maximal autocorrelation of lag 3 / weighted by Sanderson electronegativity	GETAWAY descriptors
R4e+	R maximal autocorrelation of lag 4 / weighted by Sanderson electronegativity	GETAWAY descriptors
R5e+	R maximal autocorrelation of lag 5 / weighted by Sanderson electronegativity	GETAWAY descriptors
R6e+	R maximal autocorrelation of lag 6 / weighted by Sanderson electronegativity	GETAWAY descriptors
R7e+	R maximal autocorrelation of lag 7 / weighted by Sanderson electronegativity	GETAWAY descriptors
R8e+	R maximal autocorrelation of lag 8 / weighted by Sanderson electronegativity	GETAWAY descriptors
RTe+	R maximal index / weighted by Sanderson electronegativity	GETAWAY descriptors
R1p	R autocorrelation of lag 1 / weighted by polarizability	GETAWAY descriptors
R2p	R autocorrelation of lag 2 / weighted by polarizability	GETAWAY descriptors
R3p	R autocorrelation of lag 3 / weighted by polarizability	GETAWAY descriptors
R4p	R autocorrelation of lag 4 / weighted by polarizability	GETAWAY descriptors
R5p	R autocorrelation of lag 5 / weighted by polarizability	GETAWAY descriptors
R6p	R autocorrelation of lag 6 / weighted by polarizability	GETAWAY descriptors
R7p	R autocorrelation of lag 7 / weighted by polarizability	GETAWAY descriptors
R8p	R autocorrelation of lag 8 / weighted by polarizability	GETAWAY descriptors
RTp	R total index / weighted by polarizability	GETAWAY descriptors
R1p+	R maximal autocorrelation of lag 1 / weighted by polarizability	GETAWAY descriptors
R2p+	R maximal autocorrelation of lag 2 / weighted by polarizability	GETAWAY descriptors
R3p+	R maximal autocorrelation of lag 3 / weighted by polarizability	GETAWAY descriptors
R4p+	R maximal autocorrelation of lag 4 / weighted by polarizability	GETAWAY descriptors
R5p+	R maximal autocorrelation of lag 5 / weighted by polarizability	GETAWAY descriptors
R6p+	R maximal autocorrelation of lag 6 / weighted by polarizability	GETAWAY descriptors
R7p+	R maximal autocorrelation of lag 7 / weighted by polarizability	GETAWAY descriptors
R8p+	R maximal autocorrelation of lag 8 / weighted by polarizability	GETAWAY descriptors
RTp+	R maximal index / weighted by polarizability	GETAWAY descriptors
R1i	R autocorrelation of lag 1 / weighted by ionization potential	GETAWAY descriptors
R2i	R autocorrelation of lag 2 / weighted by ionization potential	GETAWAY descriptors
R3i	R autocorrelation of lag 3 / weighted by ionization potential	GETAWAY descriptors
R4i	R autocorrelation of lag 4 / weighted by ionization potential	GETAWAY descriptors
R5i	R autocorrelation of lag 5 / weighted by ionization potential	GETAWAY descriptors
R6i	R autocorrelation of lag 6 / weighted by ionization potential	GETAWAY descriptors
R7i	R autocorrelation of lag 7 / weighted by ionization potential	GETAWAY descriptors
R8i	R autocorrelation of lag 8 / weighted by ionization potential	GETAWAY descriptors
RTi	R total index / weighted by ionization potential	GETAWAY descriptors
R1i+	R maximal autocorrelation of lag 1 / weighted by ionization potential	GETAWAY descriptors
R2i+	R maximal autocorrelation of lag 2 / weighted by ionization potential	GETAWAY descriptors
R3i+	R maximal autocorrelation of lag 3 / weighted by ionization potential	GETAWAY descriptors
R4i+	R maximal autocorrelation of lag 4 / weighted by ionization potential	GETAWAY descriptors
R5i+	R maximal autocorrelation of lag 5 / weighted by ionization potential	GETAWAY descriptors
R6i+	R maximal autocorrelation of lag 6 / weighted by ionization potential	GETAWAY descriptors

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
R7i+	R maximal autocorrelation of lag 7 / weighted by ionization potential	GETAWAY descriptors
R8i+	R maximal autocorrelation of lag 8 / weighted by ionization potential	GETAWAY descriptors
RTi+	R maximal index / weighted by ionization potential	GETAWAY descriptors
R1s	R autocorrelation of lag 1 / weighted by I-state	GETAWAY descriptors
R2s	R autocorrelation of lag 2 / weighted by I-state	GETAWAY descriptors
R3s	R autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
R4s	R autocorrelation of lag 4 / weighted by I-state	GETAWAY descriptors
R5s	R autocorrelation of lag 5 / weighted by I-state	GETAWAY descriptors
R6s	R autocorrelation of lag 6 / weighted by I-state	GETAWAY descriptors
R7s	R autocorrelation of lag 7 / weighted by I-state	GETAWAY descriptors
R8s	R autocorrelation of lag 8 / weighted by I-state	GETAWAY descriptors
RTs	R total index / weighted by I-state	GETAWAY descriptors
R1s+	R maximal autocorrelation of lag 1 / weighted by I-state	GETAWAY descriptors
R2s+	R maximal autocorrelation of lag 2 / weighted by I-state	GETAWAY descriptors
R3s+	R maximal autocorrelation of lag 3 / weighted by I-state	GETAWAY descriptors
R4s+	R maximal autocorrelation of lag 4 / weighted by I-state	GETAWAY descriptors
R5s+	R maximal autocorrelation of lag 5 / weighted by I-state	GETAWAY descriptors
R6s+	R maximal autocorrelation of lag 6 / weighted by I-state	GETAWAY descriptors
R7s+	R maximal autocorrelation of lag 7 / weighted by I-state	GETAWAY descriptors
R8s+	R maximal autocorrelation of lag 8 / weighted by I-state	GETAWAY descriptors
RTs+	R maximal index / weighted by I-state	GETAWAY descriptors
DP01	Molecular profile no. 1	Randic molecular profiles
DP02	Molecular profile no. 2	Randic molecular profiles
DP03	Molecular profile no. 3	Randic molecular profiles
DP04	Molecular profile no. 4	Randic molecular profiles
DP05	Molecular profile no. 5	Randic molecular profiles
DP06	Molecular profile no. 6	Randic molecular profiles
DP07	Molecular profile no. 7	Randic molecular profiles
DP08	Molecular profile no. 8	Randic molecular profiles
DP09	Molecular profile no. 9	Randic molecular profiles
DP10	Molecular profile no. 10	Randic molecular profiles
DP11	Molecular profile no. 11	Randic molecular profiles
DP12	Molecular profile no. 12	Randic molecular profiles
DP13	Molecular profile no. 13	Randic molecular profiles
DP14	Molecular profile no. 14	Randic molecular profiles
DP15	Molecular profile no. 15	Randic molecular profiles
DP16	Molecular profile no. 16	Randic molecular profiles
DP17	Molecular profile no. 17	Randic molecular profiles
DP18	Molecular profile no. 18	Randic molecular profiles
DP19	Molecular profile no. 19	Randic molecular profiles
DP20	Molecular profile no. 20	Randic molecular profiles
SP01	Shape profile no. 1	Randic molecular profiles
SP02	Shape profile no. 2	Randic molecular profiles
SP03	Shape profile no. 3	Randic molecular profiles
SP04	Shape profile no. 4	Randic molecular profiles
SP05	Shape profile no. 5	Randic molecular profiles
SP06	Shape profile no. 6	Randic molecular profiles
SP07	Shape profile no. 7	Randic molecular profiles
SP08	Shape profile no. 8	Randic molecular profiles
SP09	Shape profile no. 9	Randic molecular profiles
SP10	Shape profile no. 10	Randic molecular profiles
SP11	Shape profile no. 11	Randic molecular profiles
SP12	Shape profile no. 12	Randic molecular profiles
SP13	Shape profile no. 13	Randic molecular profiles
SP14	Shape profile no. 14	Randic molecular profiles
SP15	Shape profile no. 15	Randic molecular profiles

Table C.1. Continued.

Descriptor	Chemical meaning	Type*
SP16	Shape profile no. 16	Randic molecular profiles
SP17	Shape profile no. 17	Randic molecular profiles
SP18	Shape profile no. 18	Randic molecular profiles
SP19	Shape profile no. 19	Randic molecular profiles
SP20	Shape profile no. 20	Randic molecular profiles
SHP2	Average shape profile index of order 2	Randic molecular profiles
nHBonds	Number of intramolecular H-bonds (with N,O,F)	Functional group counts
G(N..N)	Sum of geometrical distances between N..N	3D Atom Pairs
G(N..O)	Sum of geometrical distances between N..O	3D Atom Pairs
G(N..S)	Sum of geometrical distances between N..S	3D Atom Pairs
G(N..F)	Sum of geometrical distances between N..F	3D Atom Pairs
G(N..Cl)	Sum of geometrical distances between N..Cl	3D Atom Pairs
G(O..O)	Sum of geometrical distances between O..O	3D Atom Pairs
G(O..S)	Sum of geometrical distances between O..S	3D Atom Pairs
G(O..F)	Sum of geometrical distances between O..F	3D Atom Pairs
G(O..Cl)	Sum of geometrical distances between O..Cl	3D Atom Pairs
G(S..Cl)	Sum of geometrical distances between S..Cl	3D Atom Pairs
G(Cl..Cl)	Sum of geometrical distances between Cl..Cl	3D Atom Pairs
CPK Volume (Å³)	CPK Volume (Å³)	Quantum chemical
CPK Area (Å²)	CPK Area (Å²)	Quantum chemical
Dipole (debye)	Dipole (debye)	Quantum chemical
E HOMO (eV)	The highest occupied molecular orbital energy (eV)	Quantum chemical
E LUMO (eV)	The lowest unoccupied molecular orbital energy (eV)	Quantum chemical
E (eV)	Gas phase energy (eV)	Quantum chemical
Gap= Elumo - Ehomo	Gap	Quantum chemical
Hardness=Gap/2	Hardness	Quantum chemical
En=-(Ehomo+Elumo)/2	Electronegativity (eV)	Quantum chemical
Z=CPK Volume (Å³) / CPK Area (Å²)	CPK Volume-CPK Area ratio (Å)	Quantum chemical
Softness= 1/Hardness	Softness	Quantum chemical
Ep=(En^2)/(2*Hardness)	Electrophilicity index	Quantum chemical

*RDF: radial distribution function, 3D-MoRSE: 3D-molecule representation of structures based on electron diffraction, WHIM: weighted holistic invariant molecular, GETAWAY: geometry, topology, and atom-weights assembly.

Table C.2. Pearson correlation results of descriptors significantly affected ($r < 0.75$).

Descriptor	PM6_vs_HF	Descriptor	HF_vs_DFT	Descriptor	PM6_vs_DFT
Mor30s	0.09673	Mor32s	0.27199	Mor24i	0.26819
Mor24i	0.11648	Mor30s	0.37997	Mor24u	0.28796
Mor24u	0.13253	Mor26s	0.39365	Mor30m	0.30933
Mor26s	0.15124	Mor27s	0.41216	Mor24e	0.32761
Mor26i	0.18151	G1u	0.44013	Mor26s	0.33790
Mor24e	0.20716	G2u	0.46833	G2u	0.34385
Mor26e	0.20908	G2i	0.47380	Mor27s	0.34551
Mor26u	0.23117	G2m	0.47665	Mor27i	0.36884
Mor30m	0.23885	Mor25s	0.49144	G2v	0.39252
Mor32i	0.24949	Mor24s	0.50115	E (eV)	0.40812
Mor32u	0.28490	HATS1i	0.50882	Mor32s	0.42116
G1u	0.28589	Mor17s	0.51916	Mor30s	0.43071
Mor30i	0.32487	HATS1u	0.52540	Mor16i	0.43177
Mor30p	0.32497	Mor21s	0.53056	Mor27e	0.43440
Mor30e	0.32753	Mor28s	0.54870	G2s	0.44967
Mor27s	0.33703	HATS1e	0.56127	Mor22s	0.45060
Mor30u	0.33745	G2e	0.56655	Mor27u	0.45075
Mor32e	0.35576	Mor23s	0.56725	G2p	0.45187
Mor30v	0.35793	Mor29s	0.56774	G1u	0.46459
G2v	0.36688	Mor31s	0.57373	Mor17s	0.47983
Mor16i	0.37997	Mor30i	0.58591	G2e	0.49207
Mor28s	0.39475	HATSi	0.58771	Mor16u	0.49820
E (eV)	0.40710	G2p	0.59208	G2i	0.50737
Mor16s	0.41072	H0i	0.59313	Mor26e	0.51619
Mor27i	0.41884	Mor24i	0.60234	Mor26i	0.52159
Mor10s	0.42563	Mor16s	0.60441	Mor30v	0.52426
G2p	0.42730	Mor30u	0.60886	Mor32i	0.52795
Mor24p	0.42831	Mor32i	0.60892	G2m	0.53222
Mor24s	0.43503	Mor30e	0.61330	Mor16e	0.54190
Mor31s	0.44289	Mor22s	0.62352	Mor32u	0.54999
Mor16u	0.46113	G1s	0.62878	Mor26u	0.55422
Mor17s	0.46946	Mor24e	0.63384	Mor24s	0.55448
Mor10i	0.48044	G2v	0.63826	G1m	0.55945
Mor17m	0.48098	Mor32u	0.64026	Mor32e	0.57606
Mor16e	0.48594	Mor24u	0.64275	G1v	0.57890
G2e	0.49234	Mor32e	0.64824	Mor24p	0.59013
Mor10e	0.49461	G2s	0.66429	Mor28s	0.59230
HATS1i	0.49580	G1m	0.66450	ISH	0.59787

Table C.2. Continued.

Descriptor	PM6_vs_HF	Descriptor	HF_vs_DFT	Descriptor	PM6_vs_DFT
Mor27e	0.49617	Mor32m	0.67126	G1i	0.62130
Mor24v	0.49931	R1u+	0.67886	Mor21s	0.62431
G2s	0.50698	Mor30m	0.68297	Mor08e	0.62527
HATS1u	0.50834	Mor30p	0.68750	Mor25s	0.62696
Mor22s	0.50854	Mor31m	0.68832	Mor08u	0.63068
Mor32m	0.51050	G1e	0.69522	Mor19s	0.63380
Mor21s	0.51425	G1p	0.69859	G1s	0.63487
Mor27u	0.51432	Mor23m	0.70150	Mor08i	0.63702
G2u	0.51459	R1i+	0.70244	G1p	0.64664
Mor32s	0.53693	HATSe	0.70516	Mor22i	0.65247
Mor24m	0.54163	Mor26i	0.71039	Mor23s	0.65614
Mor10u	0.54384	G1i	0.71465	Mor10i	0.65625
HATS1e	0.54914	R1e+	0.72512	Mor28m	0.66229
Mor32p	0.55055	Mor27i	0.72636	Mor30p	0.66354
Mor29s	0.55108	RTu+	0.72749	Mor32m	0.66570
Mor19s	0.55155	Mor26e	0.72863	Mor10e	0.66714
Mor28m	0.55839	Mor25m	0.73541	Mor17m	0.67060
H0i	0.55884	Mor26u	0.73758	Mor24v	0.67325
HATSi	0.56307	Mor17m	0.74034	Mor22e	0.67450
G2i	0.56748	RTe+	0.74221	Mor29s	0.67773
Mor22i	0.57363	Mor28m	0.74434	Mor22u	0.67923
ISH	0.57411			Mor10u	0.68900
Mor20s	0.59476			Mor16s	0.69759
Mor08e	0.60009			Mor15s	0.69820
Mor22e	0.60472			Mor24m	0.70606
Mor23m	0.60520			Mor10s	0.71006
Mor22u	0.60558			Mor19m	0.71180
Mor17v	0.60872			Softness= 1/Hardness	0.71415
Mor32v	0.61032			Mor32p	0.71876
G1v	0.63229			G1e	0.71913
Mor17i	0.63269			Mor29m	0.72301
Mor08s	0.63405			Gap= Elumo - Ehomo	0.72386
Mor28i	0.63601			Hardness=Gap/2	0.72408
R1u+	0.63703			Mor31s	0.72899
Mor31m	0.63942			Mor21m	0.73055
Mor08i	0.63946			Mor27m	0.73406
Mor23s	0.64962			Mor20s	0.73710
Mor26p	0.64962			Mor23m	0.74011

Table C.2. Continued.

Descriptor	PM6_vs_HF	Descriptor	PM6_vs_HF	Descriptor	PM6_vs_DFT
Mor17e	0.65184	G1i	0.74126	Mor13s	0.74162
Mor08u	0.65280	Mor23u	0.74170	Mor22m	0.74209
RTu+	0.66200	Mor21m	0.74186	Mor09i	0.74345
G1p	0.66345	Mor22m	0.74440	Mor08s	0.74956
Mor28e	0.66513	Mor16m	0.74813		
Mor26v	0.66821				
R1i+	0.66904				
R1e+	0.67263				
G2m	0.67573				
HATSe	0.67647				
Mor11s	0.67678				
Mor18s	0.68038				
RTe+	0.68087				
Mor17p	0.68475				
G1e	0.68631				
Mor17u	0.68667				
Mor25s	0.68755				
Mor19m	0.68934				
Mor28u	0.69154				
RTi+	0.69479				
Mor23v	0.69977				
Mor27m	0.70013				
Mor29m	0.70210				
Mor23i	0.70334				
G1m	0.70390				
Mor23e	0.70548				
Mor20i	0.70735				
G1s	0.70895				
Softness= 1/Hardness	0.71162				
E2p	0.71463				
Mor22p	0.72540				
Mor20u	0.72815				
Mor13s	0.73037				
Mor26m	0.73315				
Mor22v	0.73357				
Mor15s	0.73808				
Mor20e	0.73993				
HATS1s	0.74009				

APPENDIX D: DETAILED RESULTS OF RTL-W1 MODEL

Table D.1. RTL-W1 external set chemicals, predicted pEC_{50}/calculated pLC_{50}, model descriptors.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
110-82-7	Cyclohexane	-2.02	-2.09	0	-10.777
108-87-2	Methylcyclohexane	-1.95	-2.01	0	-10.620
1678-91-7	Ethylcyclohexane	-1.93	-1.99	0	-10.571
109-69-3	1-chlorobutane	-1.95	-2.01	0	-10.609
75-09-2	Dichloromethane	-2.09	-2.15	0	-10.921
75-34-3	1,1-dichloroethane	-2.08	-2.15	0	-10.910
107-06-2	1,2-dichloroethane	-2.11	-2.18	0	-10.975
78-87-5	1,2-dichloropropane	-2.07	-2.14	0	-10.875
142-28-9	1,3-dichloropropane	-1.99	-2.05	0	-10.699
67-66-3	Trichloromethane	-2.26	-2.33	0	-11.294
79-00-5	1,1,2-trichloroethane	-2.18	-2.25	0	-11.124
96-18-4	1,2,3-trichloropropane	-2.07	-2.14	0	-10.880
56-23-5	Carbon tetrachloride	-2.37	-2.44	0	-11.544
79-34-5	1,1,2,2-tetrachloroethane	-2.19	-2.26	0	-11.138
76-01-7	Pentachloroethane	-2.21	-2.29	0	-11.197
109-64-8	1,3-dibromopropane	-1.92	-1.98	0	-10.546
75-27-4	Bromodichloromethane	-2.17	-2.24	0	-11.097
124-48-1	Dibromochloromethane	-2.04	-2.11	0	-10.823
96-12-8	1,2-Dibromo-3-chloropropane	-1.98	-2.04	0	-10.676
542-75-6	1,3-dichloropropene	-1.72	-1.77	0	-10.097
760-23-6	3,4-Dichlorobut-1-ene	-1.96	-2.02	0	-10.632
79-01-6	Trichloroethylene	-1.61	-1.66	0	-9.860
127-18-4	Tetrachloroethene	-1.54	-1.59	0	-9.724
78-79-5	Isoprene	-1.39	-1.43	0	-9.369
111-78-4	1,5-cyclooctadiene	-1.38	-1.43	0	-9.363
3048-65-5	3a,4,7,7a-Tetrahydro-1H-indene	-1.43	-1.47	0	-9.460
16219-75-3	5-Ethylidene-8,9,10-trinorborn-2-ene	-1.33	-1.37	0	-9.252
71-36-3	1-butanol	-1.81	-1.86	0	-10.303
78-83-1	Iso-Butanol	-1.81	-1.87	0	-10.312
75-65-0	2-Methyl-2-propanol	-1.87	-1.93	0	-10.449
71-41-0	1-pentanol	-1.81	-1.86	0	-10.304
584-02-1	3-pentanol	-1.82	-1.87	0	-10.318
111-27-3	Hexanol	-1.81	-1.86	0	-10.300
111-70-6	1-heptanol	-1.81	-1.86	0	-10.305
111-87-5	1-octanol	-1.81	-1.87	0	-10.306
143-08-8	1-nonanol	-1.81	-1.87	0	-10.308
112-30-1	1-decanol	-1.81	-1.87	0	-10.310
25339-17-7	Isodecyl alcohol	-1.77	-1.83	0	-10.216
108-93-0	Cyclohexanol	-1.78	-1.84	0	-10.247
96-23-1	1,3-Dichloro-2-propanol	-2.03	-2.09	0	-10.793
107-21-1	Ethylene glycol	-1.87	-1.93	0	-10.430
80-04-6	Hydrogenatedbisphenol A	-1.78	-1.84	0	-10.249
109-86-4	2-methoxyethanol	-1.66	-1.71	0	-9.983
110-80-5	2-ethoxyethanol	-1.65	-1.70	0	-9.951
109-59-1	2-isopropoxyethanol	-1.64	-1.69	0	-9.933
111-76-2	2-butoxyethanol	-1.63	-1.68	0	-9.895
111-90-0	2-(2 Ethoxyethoxy)ethanol	-1.61	-1.66	0	-9.868
112-34-5	2-(2-Butoxyethoxy)ethanol	-1.61	-1.66	0	-9.860
60-29-7	Diethylether	-1.58	-1.63	0	-9.792

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
142-96-1	1,1'-oxybis-butane	-1.53	-1.58	0	-9.689
111-44-4	Bis(2-chloroethyl) ether	-1.75	-1.81	0	-10.183
127-90-2	2,3,3,3,2',3',3',3'-Octachlorodipropyl ether	-2.02	-2.09	0	-10.780
75-07-0	Acetaldehyde	-1.78	-1.84	0	-10.241
123-15-9	2-methylvaleraldehyde	-1.62	-1.67	0	-9.892
170-30-3	Crotonaldehyde	-1.69	-1.74	0	-10.037
111-30-8	Glutaraldehyde	-1.63	-1.68	0	-9.911
67-64-1	Acetone	-1.79	-1.85	0	-10.260
693-54-9	2-decanone	-1.73	-1.78	0	-10.117
112-12-9	2-undecanone	-1.73	-1.78	0	-10.116
593-08-8	2-tridecanone	-1.73	-1.78	0	-10.120
108-94-1	Cyclohexanone	-1.63	-1.68	0	-9.899
1502-22-3	2-(1'-Cyclohexenyl)cyclohexanone	-1.31	-1.36	0	-9.207
78-59-1	3,5,5-Trimethyl-2-cyclohexen-1-one	-1.56	-1.61	0	-9.756
141-78-6	Ethylacetate	-2.10	-2.16	0	-10.938
110-19-0	Isobutyl acetate	-2.06	-2.13	0	-10.861
111-82-0	Methyl dodecanoate	-2.07	-2.14	0	-10.879
515-84-4	Ethyl trichloroacetate	-2.15	-2.22	0	-11.055
105-53-3	Diethyl malonate	-2.20	-2.27	0	-11.157
96-33-3	Methyl acrylate	-2.12	-2.19	0	-11.000
140-88-5	Ethylacrylate	-2.07	-2.14	0	-10.878
141-32-2	N-Butyl acrylate	-2.07	-2.14	0	-10.877
818-61-1	2-hydroxyethyl acrylate	-1.87	-1.93	0	-10.444
97-88-1	N-butyl methacrylate	-1.95	-2.01	0	-10.616
688-84-6	2-Ethylhexyl methacrylate	-1.92	-1.98	0	-10.554
868-77-9	2-hydroxyethyl methacrylate	-1.87	-1.93	0	-10.444
2867-47-2	2-(Dimethylamino)ethyl methacrylate <DMMA>	-1.07	-1.11	0	-8.677
13048-33-4	Hexamethylene diacrylate	-1.94	-2.00	0	-10.587
108-05-4	Vinyl acetate	-1.71	-1.76	0	-10.082
600-07-7	2-Methylbutanoic acid	-3.23	-3.33	1	-11.009
503-74-2	3-Methylbutanoic acid	-3.32	-3.43	1	-11.222
75-98-9	Pivalic acid	-3.16	-3.26	1	-10.870
88-09-5	2-ethyl-Butanoic acid	-3.21	-3.32	1	-10.978
111-14-8	Heptanoic acid	-3.33	-3.44	1	-11.236
124-07-2	Octanoic acid	-3.32	-3.42	1	-11.212
334-48-5	Decanoic acid	-3.14	-3.25	1	-10.829
79-11-8	Chloroacetic acid	-3.28	-3.39	1	-11.134
335-67-1	Perfluorooctanoic acid	-3.81	-3.93	1	-12.291
298-12-4	Glyoxylic acid	-3.10	-3.20	1	-10.744
3821-81-6	A-fluoro-β-alanine	-2.74	-2.83	1	-9.939
79-10-7	Acrylic acid	-3.38	-3.48	1	-11.335
79-41-4	Methacrylic acid	-3.14	-3.24	1	-10.816
110-44-1	Sorbic acid	-2.70	-2.79	1	-9.853
144-62-7	Oxalic acid	-4.51	-4.65	2	-11.420
124-04-9	Adipic acid	-4.42	-4.57	2	-11.227
108-91-8	Cyclohexylamine	-1.35	-1.39	0	-9.291
141-43-5	Monoethanolamine	-1.47	-1.51	0	-9.555
115-70-8	2-Amino-2-ethylpropanediol	-1.50	-1.55	0	-9.625
109-89-7	Diethylamine	-1.17	-1.21	0	-8.885
111-42-2	Diethanolamine/2,2'-iminodiethanol	-1.31	-1.36	0	-9.210
121-44-8	Triethylamine	-0.90	-0.93	0	-8.297
102-81-8	2-(Dibutylamino)ethanol	-1.06	-1.10	0	-8.660
124-09-4	1,6-hexanediamine	-1.40	-1.44	0	-9.402
6864-37-5	2,2'-Dimethyl-4,4'-methylenebis(cyclohexylamine)	-1.33	-1.37	0	-9.255
111-18-2	N,n,n',n'-tetramethylhexamethylenediamine	-1.00	-1.03	0	-8.522
3030-47-5	N-Methyl-N,N-bis(2-dimethylaminoethyl)amine	-1.01	-1.04	0	-8.530

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
629-40-3	Octanedinitrile	-2.47	-2.55	0	-11.764
107-13-1	2-propenenitrile	-2.24	-2.31	0	-11.246
126-98-7	Methacrylonitrile	-1.99	-2.06	0	-10.707
920-37-6	2-Propenenitrile, 2-chloro-	-2.01	-2.07	0	-10.747
406-86-0	4,4,4-trifluorocrotonitrile	-2.61	-2.70	0	-12.076
1855-63-6	1-Cyclohexene-1-carbonitrile	-1.68	-1.74	0	-10.030
1118-61-2	3-amino-2-Butenenitrile	-1.31	-1.36	0	-9.205
764-42-1	2-butenedinitrile, (e)-	-2.42	-2.50	0	-11.652
75-91-2	Tert-Butylhydroperoxide	-1.55	-1.60	0	-9.736
3006-82-4	Tert-Butyl 2-ethylperoxyhexanoate	-1.58	-1.63	0	-9.799
76-06-2	Trichloronitromethane	-2.31	-2.39	0	-11.417
96-29-7	2-Butanone oxime	-1.52	-1.57	0	-9.660
100-64-1	Cyclohexanone oxime	-1.50	-1.55	0	-9.617
60-34-4	Methylhydrazine	-1.02	-1.06	0	-8.574
57-14-7	N,n-dimethylhydrazine	-1.19	-1.22	0	-8.929
657-24-9	Metformin	-1.25	-1.29	0	-9.076
110-91-8	Morpholine	-1.28	-1.32	0	-9.131
2403-88-5	2,2,6,6-Tetramethylpiperidin-4-ol	-0.93	-0.96	0	-8.360
110-85-0	Piperazine	-1.14	-1.17	0	-8.824
108-80-5	Isocyanuric acid	-2.45	-2.52	0	-11.705
470-82-6	2-Oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-	-1.39	-1.43	0	-9.368
15045-43-9	2,2,5,5-tetramethyltetrahydrofuran	-1.46	-1.51	0	-9.538
62571-86-2	Captopril	-2.24	-2.31	1	-8.833
674-82-8	But-3-en-3-olide	-1.87	-1.93	0	-10.449
106-91-2	Glycidyl methacrylate	-1.84	-1.90	0	-10.367
75-08-1	Ethanethiol	-1.19	-1.22	0	-8.929
110-66-7	Pentane-1-thiol	-1.19	-1.23	0	-8.945
111-88-6	1-Mercaptooctane <n-Octylmercaptan>	-1.20	-1.23	0	-8.950
143-10-2	1-decanethiol	-1.20	-1.23	0	-8.951
60-24-2	2-mercaptoethanol	-1.18	-1.21	0	-8.914
624-92-0	Dimethyl disulphide	-1.17	-1.21	0	-8.890
110-81-6	Diethyl disulfide	-1.18	-1.21	0	-8.914
3268-49-3	3-(Methylthio)propionaldehyde	-1.17	-1.21	0	-8.886
111-17-1	3,3'-Thiodipropionic acid	-3.39	-3.50	2	-8.948
556-61-6	Methyl isothiocyanate	-1.21	-1.25	0	-8.995
79-19-6	Thiosemicarbazide	-1.18	-1.21	0	-8.915
4189-44-0	Thiourea dioxide	-0.87	-0.90	0	-8.231
1763-23-1	Perfluorooctane sulfonic acid	-2.88	-2.98	0	-12.680
115-96-8	Tris(2-chloroethyl) phosphate	-1.97	-2.04	0	-10.673
71-43-2	Benzene	-1.51	-1.56	0	-9.638
108-88-3	Toluene	-1.33	-1.37	0	-9.237
100-41-4	Ethylbenzene	-1.34	-1.39	0	-9.275
95-47-6	O-Xylene	-1.23	-1.27	0	-9.018
108-38-3	M-Xylene	-1.25	-1.28	0	-9.057
106-42-3	P-Xylene	-1.16	-1.20	0	-8.880
103-65-1	N-Propylbenzene	-1.34	-1.38	0	-9.268
98-82-8	Isopropylbenzene	-1.34	-1.39	0	-9.275
104-51-8	Butylbenzene	-1.13	-1.16	0	-8.801
99-87-6	P-Cymene	-1.18	-1.22	0	-8.921
98-51-1	4-tert-Butyltoluene	-1.17	-1.21	0	-8.897
25321-09-9	Diisopropylbenzene	-1.25	-1.29	0	-9.076
827-52-1	Cyclohexylbenzene	-1.34	-1.38	0	-9.257
108-90-7	Chlorobenzene	-1.48	-1.52	0	-9.569
95-49-8	2-chlorotoluene	-1.38	-1.42	0	-9.348
108-41-8	3-chlorotoluene	-1.42	-1.46	0	-9.437
106-43-4	4-chlorotoluene	-1.33	-1.37	0	-9.236

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
95-50-1	1,2-dichlorobenzene	-1.47	-1.52	0	-9.564
541-73-1	1,3-dichlorobenzene	-1.56	-1.61	0	-9.754
106-46-7	1,4-dichlorobenzene	-1.47	-1.51	0	-9.546
95-73-8	2,4-dichlorotoluene	-1.41	-1.46	0	-9.428
95-75-0	3,4-dichlorotoluene	-1.37	-1.41	0	-9.331
19398-61-9	2,5-dichlorotoluene	-1.40	-1.44	0	-9.401
118-69-4	2,6-dichlorotoluene	-1.51	-1.56	0	-9.647
87-61-6	1,2,3-trichlorobenzene	-1.55	-1.60	0	-9.735
120-82-1	1,2,4-trichlorobenzene	-1.50	-1.55	0	-9.620
108-86-1	Bromobenzene	-1.51	-1.56	0	-9.647
348-61-8	4-Bromo-1,2-difluorobenzene	-1.66	-1.71	0	-9.979
98-87-3	Alpha,alpha-Dichlorotoluene	-1.64	-1.70	0	-9.939
611-19-8	2-Chlorobenzyl chloride	-1.54	-1.58	0	-9.700
98-08-8	Benzotrifluoride ((trifluoromethyl)benzene)	-1.81	-1.86	0	-10.299
402-31-3	Metaxylene hexafluoride	-2.10	-2.17	0	-10.953
98-83-9	2-phenylpropene	-1.41	-1.45	0	-9.417
1321-74-0	Divinylbenzene	-1.28	-1.32	0	-9.136
100-51-6	Benzyl alcohol	-1.50	-1.55	0	-9.622
2100-42-7	2-chlorohydroquinonedimethylether	-0.95	-0.98	0	-8.396
93-15-2	4-Allyl-1,2-dimethoxybenzene	-0.87	-0.90	0	-8.240
122-57-6	Benzalacetone	-1.47	-1.52	0	-9.560
100-52-7	Benzaldehyde	-1.70	-1.76	0	-10.065
487-68-3	2,4,6-trimethylbenzaldehyde	-1.42	-1.47	0	-9.447
123-11-5	P-Methoxybenzaldehyde	-1.36	-1.40	0	-9.314
90-02-8	Salicylaldehyde/2-hydroxybenzaldehyde	-1.41	-1.45	0	-9.425
98-86-2	Acetophenone	-1.67	-1.72	0	-9.988
84-66-2	Diethyl phthalate	-1.83	-1.89	0	-10.350
84-69-5	Diisobutyl phthalate	-1.83	-1.89	0	-10.347
84-74-2	Dibutyl phthalate	-1.83	-1.89	0	-10.364
131-17-9	Diallyl phthalate	-1.83	-1.89	0	-10.353
65-85-0	Benzoic acid	-1.76	-1.81	0	-10.191
99-04-7	M-Toluic acid	-1.56	-1.61	0	-9.759
99-94-5	4-Methylbenzoic acid	-1.63	-1.68	0	-9.897
50-78-2	Acetylsalicylic acid	-1.64	-1.69	0	-9.927
69-72-7	Salicylic acid	-1.47	-1.52	0	-9.560
99-96-7	4-Hydroxybenzoic acid	-1.55	-1.60	0	-9.740
19715-19-6	3,5-Di-tert-butylsalicylic acid	-1.21	-1.25	0	-8.986
2840-28-0	3-Amino-4-chlorobenzoic acid	-1.21	-1.25	0	-8.985
94-74-6	2-methyl-4-chlorophenoxyacetic acid	-2.34	-2.41	1	-9.048
882-09-7	Clofibric acid	-2.41	-2.48	1	-9.198
140-10-3	Trans-cinnamic acid	-2.65	-2.74	1	-9.742
108-95-2	Phenol	-1.26	-1.30	0	-9.098
106-44-5	4-cresol	-1.11	-1.15	0	-8.769
108-39-4	3-methylphenol	-1.21	-1.25	0	-8.981
95-48-7	2-methylphenol	-1.17	-1.21	0	-8.895
90-00-6	2-ethylphenol	-1.17	-1.21	0	-8.902
123-07-9	4-ethylphenol	-1.13	-1.16	0	-8.804
620-17-7	3-ethylphenol	-1.21	-1.25	0	-8.976
526-75-0	2,3-dimethylphenol	-1.14	-1.18	0	-8.829
576-26-1	2,6-dimethylphenol	-1.09	-1.13	0	-8.720
95-65-8	3,4-dimethylphenol	-1.06	-1.10	0	-8.660
95-87-4	2,5-dimethylphenol	-1.08	-1.11	0	-8.690
105-67-9	2,4-dimethylphenol	-1.03	-1.07	0	-8.586
108-68-9	3,5-dimethylphenol	-1.19	-1.22	0	-8.934
527-60-6	2,4,6-trimethylphenol	-0.97	-1.00	0	-8.450
697-82-5	2,3,5-trimethylphenol	-1.10	-1.13	0	-8.732

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
2416-94-6	2,3,6-trimethylphenol	-1.04	-1.07	0	-8.607
88-18-6	2-tert-butyl phenol	-1.14	-1.18	0	-8.826
89-72-5	O-sec-Butylphenol	-1.14	-1.18	0	-8.828
89-83-8	Thymol	-1.11	-1.14	0	-8.756
99-71-8	P-sec-Butylphenol	-1.12	-1.15	0	-8.778
14938-35-3	4-pentylphenol	-1.13	-1.16	0	-8.805
88-60-8	6-tert-Butyl-m-cresol	-1.10	-1.13	0	-8.727
2219-82-1	6-Tert-butyl-o-cresol	-1.05	-1.09	0	-8.637
2409-55-4	2-tert-Butyl-p-cresol	-1.01	-1.04	0	-8.529
1879-09-0	2-(1,1-Dimethylethyl)-4,6-dimethylphenol	-0.93	-0.96	0	-8.371
96-76-4	2,4-di-tert-butylphenol	-1.01	-1.05	0	-8.549
1806-26-4	4-n-octylphenol	-1.13	-1.17	0	-8.809
5510-99-6	2,6-Di-sec-butylphenol	-1.06	-1.10	0	-8.656
120-95-6	2,4-Di-tert-pentylphenol	-1.00	-1.03	0	-8.507
95-57-8	2-chlorophenol	-1.33	-1.37	0	-9.241
108-43-0	3-chlorophenol	-1.39	-1.44	0	-9.394
106-48-9	4-chlorophenol	-1.27	-1.31	0	-9.118
59-50-7	4-Chloro-3-methylphenol	-1.21	-1.25	0	-8.981
576-24-9	2,3-dichlorophenol	-1.44	-1.48	0	-9.479
120-83-2	2,4-dichlorophenol	-1.34	-1.38	0	-9.264
583-78-8	2,5-dichlorophenol	-1.42	-1.47	0	-9.446
87-65-0	2,6-dichlorophenol	-1.37	-1.41	0	-9.333
95-77-2	3,4-dichlorophenol	-1.34	-1.38	0	-9.256
591-35-5	3,5-dichlorophenol	-1.54	-1.59	0	-9.717
15950-66-0	2,3,4-trichlorophenol	-1.38	-1.43	0	-9.363
933-78-8	2,3,5-trichlorophenol	-1.51	-1.56	0	-9.655
933-75-5	2,3,6-trichlorophenol	-1.42	-1.47	0	-9.446
95-95-4	2,4,5-trichlorophenol	-1.38	-1.43	0	-9.365
88-06-2	2,4,6-trichlorophenol	-1.38	-1.43	0	-9.361
58-90-2	2,3,4,6-tetrachlorophenol	-1.40	-1.44	0	-9.398
106-41-2	4-bromophenol	-1.31	-1.36	0	-9.211
615-58-7	2,4-dibromophenol	-1.39	-1.43	0	-9.372
118-79-6	2,4,6-tribromophenol	-1.44	-1.49	0	-9.492
1745-81-9	2-allylphenol	-1.19	-1.23	0	-8.939
4286-23-1	4-(1-Methylethenyl)phenol	-1.17	-1.21	0	-8.904
90-05-1	2-methoxyphenol	-1.05	-1.08	0	-8.618
25013-16-5	Butylated hydroxyanisole	-0.87	-0.90	0	-8.241
99-76-3	Methyl p-hydroxybenzoate	-1.50	-1.55	0	-9.620
88-75-5	2-nitrophenol	-1.66	-1.71	0	-9.967
89-64-5	4-Chloro-2-nitrophenol	-1.59	-1.64	0	-9.826
88-30-2	3-Trifluoromethyl-4-nitrophenol	-1.91	-1.97	0	-10.517
123-31-9	Hydroquinone	-1.06	-1.10	0	-8.657
121-79-9	Propyl gallate	-1.32	-1.36	0	-9.223
62-53-3	Aniline	-0.98	-1.01	0	-8.468
95-53-4	2-methylaniline	-0.91	-0.94	0	-8.333
108-44-1	3-methylaniline	-0.94	-0.97	0	-8.394
106-49-0	4-methylaniline	-0.87	-0.90	0	-8.236
578-54-1	2-ethylaniline	-0.91	-0.94	0	-8.328
587-02-0	3-ethylaniline	-0.94	-0.97	0	-8.387
589-16-2	4-ethylaniline	-0.88	-0.91	0	-8.261
87-59-2	2,3-dimethylaniline	-0.89	-0.92	0	-8.284
87-62-7	2,6-dimethylaniline	-0.87	-0.90	0	-8.227
95-64-7	3,4-dimethylaniline	-0.84	-0.86	0	-8.158
95-68-1	2,4-dimethylaniline	-0.82	-0.85	0	-8.123
95-78-3	2,5-dimethylaniline	-0.87	-0.90	0	-8.233
108-69-0	3,5-dimethylaniline	-0.92	-0.95	0	-8.337

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
579-66-8	2,6-diethylaniline	-0.86	-0.89	0	-8.214
99-88-7	4-isopropylaniline	-0.88	-0.91	0	-8.263
88-05-1	2,4,6-trimethylaniline	-0.78	-0.80	0	-8.027
95-51-2	2-chloroaniline	-1.04	-1.07	0	-8.612
108-42-9	3-chloroaniline	-1.12	-1.15	0	-8.784
106-47-8	4-chloroaniline	-1.02	-1.06	0	-8.573
95-81-8	2-Chloro-5-methylaniline	-1.00	-1.03	0	-8.508
95-76-1	3,4-dichloroaniline	-1.10	-1.14	0	-8.741
95-82-9	2,5-dichloroaniline	-1.15	-1.19	0	-8.851
554-00-7	2,4-dichloroaniline	-1.08	-1.12	0	-8.700
608-27-5	2,3-dichloroaniline	-1.14	-1.18	0	-8.826
608-31-1	2,6-dichloroaniline	-1.10	-1.14	0	-8.751
626-43-7	3,5-dichloroaniline	-1.25	-1.28	0	-9.061
634-67-3	2,3,4-trichloroaniline	-1.13	-1.17	0	-8.808
634-93-5	2,4,6-trichloroaniline	-1.14	-1.17	0	-8.822
636-30-6	2,4,5-trichloroaniline	-1.14	-1.18	0	-8.828
634-91-3	3,4,5-trichloroaniline	-1.17	-1.21	0	-8.897
104-94-9	P-Anisidine	-0.78	-0.80	0	-8.034
88-74-4	2-nitroaniline	-1.32	-1.36	0	-9.228
99-09-2	3-nitroaniline	-1.31	-1.36	0	-9.208
100-01-6	4-nitroaniline	-1.36	-1.40	0	-9.312
99-52-5	2-Methyl-4-nitroaniline	-1.29	-1.33	0	-9.146
89-63-4	4-Chloro-2-nitroaniline	-1.32	-1.36	0	-9.219
96-96-8	2-Nitro-p-anisidine	-1.05	-1.08	0	-8.626
103-69-5	N-ethylaniline	-0.75	-0.78	0	-7.973
121-69-7	N,n-dimethylaniline	-0.62	-0.64	0	-7.691
91-66-7	N,n-diethylaniline	-0.65	-0.67	0	-7.742
106-50-3	P-Phenylenediamine	-0.70	-0.72	0	-7.852
108-45-2	M-Phenylenediamine	-0.86	-0.89	0	-8.209
95-54-5	O-Phenylenediamine	-0.84	-0.86	0	-8.164
95-70-5	2,5-diaminotoluene	-0.66	-0.68	0	-7.772
95-80-7	2,4-diaminotoluene	-0.76	-0.78	0	-7.990
101-96-2	1,4-Benzenediamine, N,N'-bis(1-methylpropyl)-	-0.28	-0.28	0	-6.915
85068-29-7	3,5-Bis(trifluoromethyl)benzylamine	-1.71	-1.76	0	-10.079
1477-55-0	M-Phenylenebis(methylamine)	-1.42	-1.46	0	-9.436
29122-68-7	Atenolol	-1.15	-1.19	0	-8.848
95-55-6	2-aminophenol	-1.19	-1.22	0	-8.934
123-30-8	4-aminophenol	-0.86	-0.89	0	-8.210
591-27-5	3-aminophenol	-1.01	-1.05	0	-8.555
119-34-6	4-Amino-2-nitrophenol	-1.19	-1.22	0	-8.928
98-95-3	Nitrobenzene	-1.95	-2.01	0	-10.615
88-72-2	2-nitrotoluene	-1.75	-1.81	0	-10.184
99-99-0	4-methylnitrobenzene	-1.81	-1.87	0	-10.306
88-73-3	2-chloronitrobenzene	-1.79	-1.85	0	-10.262
121-73-3	3-chloronitrobenzene	-1.81	-1.87	0	-10.311
100-00-5	4-chloronitrobenzene	-1.88	-1.94	0	-10.472
13290-74-9	4-Chloro-3-methylnitrobenzene	-1.77	-1.82	0	-10.207
99-54-7	3,4-dichloronitrobenzene	-1.80	-1.86	0	-10.285
89-69-0	1,2,4-trichloro-5-nitrobenzene	-1.76	-1.81	0	-10.192
350-30-1	2-Chloro-1-fluoro-4-nitrobenzene	-1.87	-1.93	0	-10.427
100-14-1	Alpha-Chloro-4-nitrotoluene	-1.95	-2.01	0	-10.614
91-23-6	2-nitroanisole	-1.54	-1.58	0	-9.701
555-03-3	3-nitroanisole	-1.50	-1.55	0	-9.631
606-20-2	2,6-dinitrotoluene	-2.13	-2.20	0	-11.007
97-00-7	1-Chloro-2,4-dinitrobenzene	-2.14	-2.21	0	-11.035
534-52-1	4,6-Dinitro-o-cresol	-1.89	-1.95	0	-10.494

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
88-85-7	2-(1-methylpropyl)-4,6-dinitro-Phenol	-1.88	-1.94	0	-10.460
40487-42-1	Pendimethalin	-1.31	-1.36	0	-9.213
1582-09-8	Trifluralin	-1.66	-1.71	0	-9.968
55283-68-6	Ethalfluraline	-1.66	-1.71	0	-9.984
29091-05-2	Dinitramine	-1.54	-1.59	0	-9.722
873-32-5	O-Chlorobenzonitrile	-1.68	-1.74	0	-10.030
91-15-6	Phthalonitrile	-1.97	-2.03	0	-10.657
140-29-4	Benzyl cyanide	-1.60	-1.65	0	-9.854
23950-58-5	Propyzamide	-1.64	-1.70	0	-9.944
51218-45-2	Metolachlor	-1.44	-1.49	0	-9.489
23184-66-9	Butachlor/N-(butoxymethyl)-2-chloro-2',6'-diethylacetanilide	-1.41	-1.45	0	-9.424
51218-49-6	Pretilachlor/2-chloro-2',6'-diethyl-N-(2-propoxyethyl)acetanilide	-1.42	-1.47	0	-9.454
93-68-5	O-Acetoacetotoluidide	-1.07	-1.10	0	-8.666
57837-19-1	Metalaxyl/methyl-(2-methoxyacetyl)-N-(2,6-xylyl)-DL-alaninate	-1.41	-1.45	0	-9.418
103-90-2	Paracetamol	-0.99	-1.02	0	-8.492
3766-81-2	Fenobucarb/2-sec-butylphenyl N-methylcarbamate	-1.36	-1.40	0	-9.313
114-26-1	Propoxur	-1.10	-1.14	0	-8.753
34123-59-6	Isoproturon/1,1-dimethyl-3-(8-isopropylphenyl)-urea	-0.94	-0.97	0	-8.385
330-54-1	Diuron/1-(3,4 dichlorophenyl)-3,3 dimethyl urea	-1.14	-1.18	0	-8.830
5329 12 4	2,4,6-trichlorophenylhydrazine	-1.07	-1.10	0	-8.674
108-98-5	Benzenethiol	-1.09	-1.12	0	-8.713
28249-77-6	Thiobencarb	-1.26	-1.30	0	-9.096
88-19-7	o-toluenesulfonamide	-1.62	-1.67	0	-9.876
63-74-1	Sulphanilamide	-1.30	-1.34	0	-9.180
98-59-9	4-toluenesulfonyl chloride (p-toluene sulfonyl chloride stabilised)	-1.93	-2.00	0	-10.575
121-03-9	4-nitrotoluene-2-sulphonic acid	-2.10	-2.16	0	-10.943
15318-45-3	Thiamphenicol	-1.87	-1.93	0	-10.434
122-14-5	Fenitrothion	-1.65	-1.70	0	-9.953
26087-47-8	Iprobenfos	-1.31	-1.36	0	-9.210
110-86-1	Pyridine	-1.72	-1.77	0	-10.101
10500-57-9	5,6,7,8-tetrahydroquinoline	-1.37	-1.42	0	-9.338
100-43-6	4-vinylpyridine	-1.72	-1.77	0	-10.097
100-69-6	2-vinylpyridine	-1.47	-1.51	0	-9.552
462-08-8	3-aminopyridine	-1.14	-1.17	0	-8.824
504-24-5	4-aminopyridine	-1.29	-1.33	0	-9.148
504-29-0	2-aminopyridine	-1.17	-1.21	0	-8.894
1007-28-9	Atrazine-deisopropyl	-1.47	-1.52	0	-9.564
122-34-9	Simazine	-1.40	-1.45	0	-9.413
1912-24-9	Atrazine	-1.40	-1.45	0	-9.408
5915-41-3	Terbuthylazine	-1.39	-1.43	0	-9.379
33693-04-8	Terbumeton	-1.29	-1.33	0	-9.156
NA	2-methylthio-4-tert-butylamino-6-amino-s-triazine	-1.13	-1.16	0	-8.796
1014-70-6	Simetryn	-1.08	-1.11	0	-8.690
886-50-0	Terbutryn	-1.08	-1.11	0	-8.688
28159-98-0	Irgarol 1051/2-methylthio-4-tert-butylamino-6-cyclopropylamino-s-triazine	-1.11	-1.15	0	-8.769
51-21-8	5-fluorouracil	-1.79	-1.85	0	-10.259
21087-64-9	Metribuzin	-1.26	-1.30	0	-9.104
110-02-1	Thiophene	-1.24	-1.28	0	-9.046
443-48-1	Metronidazole	-1.77	-1.82	0	-10.213
61-82-5	3-amino-1,2,4-triazole	-1.63	-1.68	0	-9.911
92-52-4	Biphenyl	-1.29	-1.33	0	-9.154
5707-44-8	4-ethyl-1,1'-biphenyl	-1.19	-1.22	0	-8.932
90-43-7	2-phenylphenol	-1.22	-1.26	0	-9.001
92-69-3	p-phenylphenol	-1.15	-1.18	0	-8.845
92-88-6	4,4'-dihydroxy-biphenyl	-1.08	-1.11	0	-8.694
119-93-7	o-tolidine	-0.74	-0.76	0	-7.937

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
91-94-1	3,3'-dichlorobenzidine	-0.90	-0.93	0	-8.304
58-14-0	Pyrimethamine	-1.20	-1.24	0	-8.969
91-76-9	2,4-diamino-6-phenyl-s-triazine	-1.49	-1.53	0	-9.587
1698-60-8	Chloridazon/5-amino-4-chloro-2-pheny l-3(2H)-pyridazinone	-1.27	-1.31	0	-9.119
51963-82-7	Benzenamine,2,5-diethoxy-4-(4-morpholinyl)-	-0.44	-0.45	0	-7.281
32809-16-8	N-(3,5-dichlorophenyl)-1,2-dimethylcyclopropane-1,2-dicarboximide	-1.49	-1.53	0	-9.591
18854-01-8	Isoxathion	-1.43	-1.48	0	-9.472
19666-30-9	Oxadiazon	-1.15	-1.18	0	-8.835
147-94-4	Cytarabine	-1.48	-1.53	0	-9.575
95058-81-4	Gemcitabine	-1.70	-1.75	0	-10.058
96-09-3	Styrene-7,8-oxide	-1.49	-1.54	0	-9.614
901-44-0	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane	-1.10	-1.13	0	-8.727
599-64-4	4-(α,α-dimethylbenzyl)phenol	-1.15	-1.18	0	-8.837
620-92-8	4,4'-dihydroxydiphenylmethane	-1.12	-1.16	0	-8.793
80-05-7	Bisphenol A	-1.13	-1.16	0	-8.798
79-94-7	Tetrabromobisphenol A	-1.34	-1.38	0	-9.261
101-77-9	4,4'-methylenedianiline	-0.86	-0.88	0	-8.200
101-84-8	Diphenyl ether	-1.18	-1.22	0	-8.922
3380-34-5	Triclosan	-1.37	-1.42	0	-9.344
101-80-4	4,4'-diaminodiphenyl ether	-0.78	-0.80	0	-8.027
42874-03-3	Oxyfluorfen	-1.61	-1.66	0	-9.863
103-50-4	Dibenzyl ether	-1.44	-1.49	0	-9.486
119-61-9	Benzophenone	-1.65	-1.70	0	-9.948
131-57-7	2-hydroxy-4-methoxybenzophenone	-1.36	-1.41	0	-9.316
31127-54-5	2,3,4,4'-tetrahydroxybenzophenon	-1.25	-1.28	0	-9.063
122-39-4	Diphenylamine	-0.82	-0.85	0	-8.134
620-93-9	Di-p-tolylamine	-0.68	-0.70	0	-7.806
13684-63-4	Phenmedipham	-1.25	-1.28	0	-9.057
101-20-2	3,4,4'-trichlorodiphenylurea	-1.24	-1.28	0	-9.041
102-06-7	1,3-diphenylguanidine	-0.98	-1.01	0	-8.474
97-39-2	N,N'-bis(2-methylphenyl)guanidine	-1.00	-1.03	0	-8.520
122-66-7	Hydrazobenzene	-0.91	-0.94	0	-8.321
60-09-3	Aniline, p-(phenylazo)-（p-aminoazobenzene）	-0.96	-0.99	0	-8.427
80-09-1	Bis(4-hydroxyphenyl)sulfone	-1.54	-1.58	0	-9.701
30171-80-3	Dibromocresyl glycidyl ether	-1.34	-1.39	0	-9.282
738-70-5	Trimethoprim	-1.10	-1.13	0	-8.731
68-35-9	Sulfadiazine	-1.34	-1.38	0	-9.258
57-68-1	Sulfamethazine/sulfadimidine	-1.30	-1.35	0	-9.186
122-11-2	Sulfadimethoxine	-1.27	-1.31	0	-9.125
64902-72-3	Chlorsulfuron	-1.74	-1.79	0	-10.154
77732-09-3	Oxadixyl	-1.47	-1.52	0	-9.561
723-46-6	Sulfamethoxazole	-1.27	-1.31	0	-9.112
90-12-0	1-methylnaphthalene	-1.07	-1.10	0	-8.668
91-57-6	2-methylnaphthalene	-1.10	-1.13	0	-8.729
573-98-8	1,2-dimethylnaphthalene	-1.00	-1.03	0	-8.522
575-41-7	1,3-dimethylnaphthalene	-1.01	-1.04	0	-8.536
582-16-1	2,7-dimethylnaphthalene	-1.06	-1.09	0	-8.652
29253-36-9	Isopropylnaphthalene	-1.07	-1.10	0	-8.667
525-66-6	Propranolol	-1.00	-1.03	0	-8.511
135-19-3	B-naphthol	-1.10	-1.14	0	-8.747
91-59-8	B-naphthylamine	-0.88	-0.91	0	-8.252
479-27-6	1,8-naphthylenediamine	-0.76	-0.78	0	-7.987
2243-62-1	1,5-naphthalenediamine	-0.71	-0.73	0	-7.874
92-70-6	3-Hydroxy-2-naphthoic acid	-1.20	-1.24	0	-8.967
58-27-5	2-methyl-1,4-naphthoquinone	-1.82	-1.88	0	-10.334
117-80-6	2,3-dichloro-1,4-naphthoquinone	-1.73	-1.79	0	-10.133

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
1785-65-5	2-acetoxy-1,4-naphthoquinone	-1.86	-1.92	0	-10.406
83-72-7	2-hydroxy-1,4-naphthoquinone	-1.54	-1.59	0	-9.720
91-22-5	Quinoline	-1.39	-1.43	0	-9.374
91-53-2	6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	-0.51	-0.53	0	-7.442
148-24-3	8-hydroxyquinoline	-1.15	-1.19	0	-8.847
22720-75-8	1-benzo[b]thien-2-ylethan-1-one	-1.11	-1.15	0	-8.765
95-31-8	N-(tert-butyl)-2-benzothiazolylsulfenamide	-0.99	-1.02	0	-8.499
95-33-0	N-cyclohexyl-2-benzothiazolylsulfenamide	-1.02	-1.05	0	-8.558
149-30-4	2-mercaptobenzothiazole	-1.04	-1.07	0	-8.614
11070-44-3	Tetrahydromethylphthalic anhydride	-2.05	-2.11	0	-10.826
85-44-9	Phthalic anhydride	-2.14	-2.21	0	-11.037
117-08-8	Tetrachlorophthalic anhydride	-1.89	-1.95	0	-10.482
83-32-9	Acenaphthene	-1.03	-1.06	0	-8.579
85-01-8	Phenanthrene	-1.11	-1.14	0	-8.765
86-73-7	Fluorene	-1.10	-1.13	0	-8.735
193-39-5	Indeno[1,2,3-cd]pyrene	-0.92	-0.95	0	-8.335
5522-43-0	1-nitropyrene	-1.22	-1.26	0	-9.007
1484-13-5	9-vinylcarbazole	-0.88	-0.91	0	-8.246
298-46-4	Carbamazepine	-1.12	-1.15	0	-8.779
2222-33-5	Dibenzo[b,f]cyclohepten-1-one	-1.32	-1.36	0	-9.234
132-65-0	Dibenzothiophene	-0.99	-1.02	0	-8.493
1916-55-8	2-acetamidophenoxazin-3-one	-1.25	-1.29	0	-9.067
1916-59-2	2-aminophenoxazin-3-one	-1.09	-1.13	0	-8.717
NA	2-amino-7-methoxyphenoxazin-3-one	-0.99	-1.02	0	-8.496
92-84-2	Phenothiazine	-0.62	-0.64	0	-7.666
14698-29-4	Oxolinic acid	-2.30	-2.37	1	-8.960
42835-25-6	Flumequine	-2.37	-2.45	1	-9.130
2439 01 2	6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one	-1.42	-1.46	0	-9.439
90-30-2	1-(N-phenylamino)-naphthalene	-0.78	-0.80	0	-8.032
135-88-6	N-phenyl-2-naphthylamine	-0.79	-0.81	0	-8.046
88426-33-9	Buparvaquone	-1.45	-1.50	0	-9.513
79617-96-2	Sertraline (hydrochloride)	-1.23	-1.27	0	-9.023
70458-96-7	Norfloxacin	-2.25	-2.32	1	-8.850
85721-33-1	Ciprofloxacin	-2.13	-2.20	1	-8.591
93106-60-6	Enrofloxacin	-2.08	-2.14	1	-8.469
98079-51-7	Lomefloxacin	-2.13	-2.19	1	-8.585
948-65-2	2-phenylindole	-0.96	-0.99	0	-8.427
115-86-6	Triphenyl phosphate	-1.44	-1.49	0	-9.492
27955-94-8	Phenol,4,4',4"-ethylidynetris-	-1.16	-1.20	0	-8.880
60-54-8	Tetracycline	-1.45	-1.50	0	-9.520
79-57-2	Oxytetracycline	-1.29	-1.34	0	-9.172
82419-36-1	Ofloxacin	-1.99	-2.05	1	-8.272
95233-18-4	Atovaquone	-1.41	-1.45	0	-9.419
23696-28-8	Olaquindox	-1.11	-1.15	0	-8.773
58-89-9	Lindane (γ-HCH)	-2.17	-2.24	0	-11.100
75-35-4	1,1-dichloroethylene	-1.77	-1.83	0	-10.220
67-56-1	Methanol	-1.91	-1.97	0	-10.530
67-63-0	2-propanol	-1.84	-1.90	0	-10.370
20679-58-7	Acetic acid, bromo-, 2-butene-1,4-diyl ester (Fennosan F50)	-1.88	-1.94	0	-10.470
80-62-6	Methyl methacrylate	-1.97	-2.03	0	-10.650
75-64-9	t-butylamine	-1.35	-1.39	0	-9.290
124-40-3	Dimethylamine	-1.19	-1.23	0	-8.940
108-18-9	Diisopropylamine	-1.06	-1.09	0	-8.650
111-92-2	Dibutylamine	-1.14	-1.17	0	-8.820
68-12-2	Dimethylformamide	-1.39	-1.44	0	-9.390
62-75-9	Dimethylnitrosamine	-1.42	-1.47	0	-9.450

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
55-18-5	Diethylnitrosamine	-1.35	-1.40	0	-9.300
99129-21-2	Clethodim	-1.05	-1.08	0	-8.630
62-56-6	Thiourea	-1.08	-1.11	0	-8.690
2212-67-1	Molinate	-1.24	-1.28	0	-9.050
77182-82-2	Glufosinate	-2.88	-2.97	1	-10.240
1071-83-6	Glyphosate	-2.54	-2.62	1	-9.500
126-72-7	Tris-(2,3-dibromopropyl)-phoshate	-1.90	-1.96	0	-10.500
115-29-7	Endosulfan	-1.67	-1.72	0	-9.990
131-11-3	Dimethyl phthalate	-1.87	-1.93	0	-10.440
644-35-9	2-n-propylphenol	-1.16	-1.20	0	-8.880
98-54-4	P-tert-butylphenol	-1.12	-1.15	0	-8.780
104-40-5	4-n-nonylphenol	-1.13	-1.17	0	-8.810
609-19-8	3,4,5-trichlorophenol	-1.41	-1.45	0	-9.420
935-95-5	2,3,5,6-tetrachlorophenol	-1.49	-1.54	0	-9.610
4901-51-3	2,3,4,5-tetrachlorophenol	-1.43	-1.48	0	-9.470
2460-49-3	4,5-dichloroguaiacol	-1.15	-1.19	0	-8.860
2668-24-8	4,5,6-trichloroguaiacol	-1.21	-1.25	0	-8.990
57057-83-7	3,4,5-trichloroguaiacol	-1.36	-1.41	0	-9.320
2539-17-5	Tetrachloroguaiacol	-1.38	-1.43	0	-9.360
2539-26-6	3,4,5-trichloro-2,6-dimethoxyphenol	-1.19	-1.22	0	-8.930
100-02-7	4-nitrophenol	-1.73	-1.79	0	-10.130
1689-84-5	Bromoxynil	-1.60	-1.65	0	-9.840
108-46-3	Resorcinol	-1.28	-1.32	0	-9.140
120-80-9	Catechol	-1.14	-1.17	0	-8.820
615-67-8	Chlorohydroquinone	-1.15	-1.19	0	-8.860
95-88-5	4-chlororesorcinol	-1.27	-1.31	0	-9.120
2138-22-9	4-chlorocatechol	-1.20	-1.24	0	-8.960
3428-24-8	4,5-dichlorocatechol	-1.24	-1.28	0	-9.040
3978-67-4	3,4-dichlorocatechol	-1.28	-1.32	0	-9.130
13673-92-2	3,5-dichlorocatechol	-1.33	-1.37	0	-9.240
137-19-9	4,6-dichlororesorcinol	-1.32	-1.36	0	-9.220
56961-20-7	3,4,5-trichlorocatechol	-1.30	-1.35	0	-9.190
32139-72-3	3,4,6-trichlorocatechol	-1.33	-1.37	0	-9.250
1198-55-6	Tetrachlorocatechol	-1.35	-1.40	0	-9.300
87-87-6	Tetrachlorohydroquinone	-1.29	-1.33	0	-9.150
87-66-1	1,2,3-trihydroxybenzene	-1.18	-1.21	0	-8.910
99-55-8	2-amino-4-nitrotoluene	-1.26	-1.30	0	-9.090
119-32-4	4-amino-2-nitrotoluene	-1.20	-1.24	0	-8.970
603-83-8	2-amino-6-nitrotoluene	-1.24	-1.28	0	-9.040
19406-51-0	4-amino-2,6-dinitrotoluene	-1.48	-1.53	0	-9.580
35572-78-2	2-amino-4,6-dinitrotoluene	-1.54	-1.58	0	-9.700
823-40-5	2,6-diaminotoluene	-0.86	-0.89	0	-8.210
6629-29-4	2,4-diamino-6-nitrotoluene	-1.07	-1.11	0	-8.680
59229-75-3	2,6-diamino-4-nitrotoluene	-1.18	-1.21	0	-8.910
56-75-7	Chloramphenicol	-1.92	-1.99	0	-10.560
121-14-2	2,4-dinitrotoluene	-2.16	-2.23	0	-11.090
118-96-7	2,4,6-trinitrotoluene	-2.47	-2.55	0	-11.760
1194-65-6	2,6-dichlorobenzonitrile	-1.74	-1.79	0	-10.150
1897-45-6	Tetrachloroisophthalonitrile	-1.83	-1.89	0	-10.360
34256-82-1	Acetochlor	-1.43	-1.47	0	-9.460
15545-48-9	Chlorotoluron	-1.04	-1.07	0	-8.610
23564-05-8	Dimethyl 4,4'-(o-phenylente) bis(3-thioa(lophanate)	-1.24	-1.28	0	-9.040
57-67-0	Sulfaguanidine	-1.24	-1.28	0	-9.050
73231-34-2	Florfenicol	-1.86	-1.92	0	-10.420
64249-01-0	Anilofos	-1.38	-1.43	0	-9.360
22224-92-6	Fenamiphos	-1.00	-1.03	0	-8.510

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
69377-81-7	Fluroxypyr	-2.53	-2.61	1	-9.470
21725-46-2	Cyanazine	-1.48	-1.52	0	-9.570
834-12-8	Ametryn	-1.08	-1.11	0	-8.690
7287-19-6	Prometryn	-1.08	-1.11	0	-8.690
59-87-0	Nitrofurazone	-1.56	-1.61	0	-9.760
34014-18-1	Tebuthiuron	-1.30	-1.34	0	-9.180
119-12-0	Pyridaphenthion	-1.25	-1.29	0	-9.070
24096-53-5	N-(3,5-dichlorophenyl)succinidide	-1.52	-1.57	0	-9.670
36734-19-7	3-(3,5-dichlorophenyl)-N-isopropyl-2,4-dioxoimidazolidine-l-carboxamide	-1.58	-1.63	0	-9.790
39807-15-3	Oxadiargyl	-1.13	-1.17	0	-8.810
51338-27-3	Methyl 2-[4-(2,4-dichlorophenoxy)phenoxy]propanoate	-1.21	-1.25	0	-8.990
40843-25-2	Diclofop-P	-2.32	-2.40	1	-9.020
40843-73-0	4-(2,4-dichlorophenoxy)-phenol	-1.21	-1.25	0	-8.980
68359-37-5	Beta-cyfluthrin	-1.34	-1.38	0	-9.260
85-68-7	Butylbenzyl phthalate	-1.49	-1.53	0	-9.590
71626-11-4	R-(－)-benalaxyl/Rac-benalaxyl/S-(+)-benalaxyl	-1.42	-1.46	0	-9.440
126833-17-8	Fenhexamid	-1.08	-1.12	0	-8.700
72619-32-0	Haloxyfop-R	-1.56	-1.61	0	-9.760
83066-88-0	Fluazifop-p	-2.66	-2.75	1	-9.770
83055-99-6	Bensulfuron-methyl	-1.48	-1.52	0	-9.570
90982-32-4	Chlorimuron-ethyl	-1.63	-1.69	0	-9.920
111991-09-4	Nicosulfuron	-1.36	-1.40	0	-9.310
136849-15-5	Cyclosulfamuron	-1.42	-1.46	0	-9.440
74223-64-6	Metsulfuron-methyl	-1.83	-1.89	0	-10.360
106040-48-6	Tribenuron	-1.71	-1.77	0	-10.090
111353-84-5	Ethametsulfuron	-1.53	-1.58	0	-9.690
79319-85-0	N,N'-methylene-di(2-amino-5-mercapto-1,3,4-thiodiazole)	-1.12	-1.16	0	-8.790
93697-74-6	Pyrazosulfuron-ethyl	-1.57	-1.62	0	-9.770
84087-01-4	Quinclorac	-1.58	-1.63	0	-9.800
52316-55-9	Carbendazim	-1.15	-1.19	0	-8.860
17804-35-2	Methyl-l-(butylcarbamoyl)-2-benzimidazole carbamate	-1.22	-1.26	0	-9.000
18691-97-9	Methabenzthiazuron	-1.17	-1.21	0	-8.900
25059-80-7	Benazolin-ethyl	-1.34	-1.38	0	-9.270
260-94-6	Acridine	-1.15	-1.18	0	-8.840
59-40-5	Sulfaquinoxaline	-1.19	-1.22	0	-8.930
94051-08-8	Quizalofop-p	-2.51	-2.59	1	-9.420
73250-68-7	Mefenacet	-1.23	-1.27	0	-9.020
95617-09-7	Fenoxaprop	-2.39	-2.47	1	-9.170
98967-40-9	Flumetsulam	-1.22	-1.26	0	-9.000
139-91-3	Furaltadone	-1.49	-1.53	0	-9.590
87818-31-3	Cinmethylin	-1.34	-1.38	0	-9.270
125401-75-4	Bispyribac	-1.56	-1.61	0	-9.760
564-25-0	Deoxytetracycline	-1.16	-1.20	0	-8.870
57-62-5	Chlorotetracycline	-1.12	-1.16	0	-8.790
986-85-4	Levofloxacin	-2.00	-2.06	1	-8.300
41083-11-8	Azocy-clotin/1-(tricyclohexylstannyl)-1H-1,2,4-triazole	-1.38	-1.42	0	-9.350
76-87-9	Fentin hydroxide	-1.46	-1.51	0	-9.540
13121-70-5	Cyhexatin	-1.27	-1.31	0	-9.110
55268-75-2	Cefuroxime	-2.45	-2.53	1	-9.300
26787-78-0	Amoxicillin	-2.53	-2.61	1	-9.480
15686-71-2	Cephalexin	-2.47	-2.54	1	-9.330
2022-85-7	5-fluorocytosine	-1.55	-1.60	0	-9.730
16110-51-3	Cromolyn	-3.62	-3.73	2	-9.450
58-08-2	Caffeine	-1.28	-1.32	0	-9.140
73-22-3	L-tryptophan	-2.23	-2.31	1	-8.820
59-05-2	Methotrexate	-3.23	-3.33	2	-8.590

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
51-52-5	Propylthiouracil	-1.32	-1.36	0	-9.220
60-80-0	Antipyrine	-1.03	-1.06	0	-8.580
87-08-1	Phenoxymethylpenicillinic acid	-2.52	-2.60	1	-9.440
64544-07-6	Cefuroxime axetil	-1.33	-1.37	0	-9.250
33419-42-0	Etoposide	-1.01	-1.05	0	-8.550
51481-61-9	Cimetidine	-1.10	-1.13	0	-8.730
94-20-2	Chlorpropamide	-1.64	-1.70	0	-9.940
3930-20-9	Sotalol	-1.10	-1.14	0	-8.750
58-93-5	Hydrochlorothiazide	-1.59	-1.64	0	-9.830
1156-19-0	Tolazamide	-1.48	-1.52	0	-9.570
50-23-7	Hydrocortisone	-1.54	-1.59	0	-9.710
42200-33-9	Nadolol	-1.24	-1.28	0	-9.050
50-24-8	Prednisolone	-1.57	-1.62	0	-9.780
51-34-3	Scopolamine	-1.24	-1.28	0	-9.050
26839-75-8	Timolol	-0.83	-0.86	0	-8.150
37350-58-6	Metoprolol	-1.11	-1.15	0	-8.770
1091-85-6	Dansylglycine	-2.51	-2.59	1	-9.420
137-58-6	Lidocaine	-1.12	-1.15	0	-8.780
83-43-2	Methylprednisolone	-1.56	-1.61	0	-9.750
64-77-7	Tolbutamide	-1.57	-1.62	0	-9.770
526-08-9	Sulfaphenazole	-1.24	-1.28	0	-9.050
37517-30-9	Acebutolol	-1.06	-1.10	0	-8.660
59-46-1	Procaine	-1.03	-1.06	0	-8.580
63590-64-7	Terazosin	-0.81	-0.84	0	-8.110
6452-71-7	Oxprenolol	-1.29	-1.33	0	-9.150
84057-84-1	Lamotrigine	-1.29	-1.34	0	-9.170
4205-90-7	Clonidine	-1.05	-1.08	0	-8.630
13523-86-9	Pindolol	-0.92	-0.95	0	-8.340
54-31-9	Furosemide	-1.48	-1.52	0	-9.570
66357-35-5	Ranitidine	-1.11	-1.14	0	-8.760
7689 03 4	Camptothecin	-1.36	-1.40	0	-9.310
34841-39-9	Bupropion	-0.97	-1.00	0	-8.450
103628-46-2	Sumatriptan	-0.86	-0.89	0	-8.210
81-81-2	Warfarin	-1.40	-1.44	0	-9.400
28395-03-1	Bumetanide	-1.13	-1.16	0	-8.800
129-20-4	Oxyphenbutazone	-1.01	-1.04	0	-8.540
87848-99-5	Acrivastine	-2.26	-2.33	1	-8.870
57-41-0	Phenytoin	-1.61	-1.66	0	-9.860
13655-52-2	Alprenolol	-1.14	-1.17	0	-8.820
19216-56-9	Prazosin	-0.85	-0.87	0	-8.180
36894-69-6	Labetalol	-1.25	-1.28	0	-9.060
50-33-9	Phenylbutazone	-1.05	-1.08	0	-8.630
94-24-6	Tetracaine	-0.97	-1.00	0	-8.460
6990 06 3	Fusidic acid	-2.50	-2.58	1	-9.400
303-81-1	Novobiocin	-1.10	-1.13	0	-8.730
99614-02-5	Ondansetron	-1.12	-1.15	0	-8.780
548-73-2	Droperidol	-1.11	-1.15	0	-8.770
56-54-2	Quinidine	-1.08	-1.11	0	-8.690
53-86-1	Indomethacin	-2.23	-2.31	1	-8.820
130-95-0	Quinine	-1.11	-1.14	0	-8.760
599-79-1	Sulfasalazine	-1.29	-1.34	0	-9.170
57-83-0	Progesterone	-1.53	-1.58	0	-9.690
50-47-5	Desipramine	-0.72	-0.74	0	-7.900
50-28-2	Estradiol	-1.08	-1.12	0	-8.700
10238-21-8	Glibenclamide	-1.34	-1.38	0	-9.260
58-22-0	Testosterone	-1.55	-1.60	0	-9.740

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
50-49-7	Imipramine	-0.64	-0.66	0	-7.720
65277-42-1	Ketoconazole	-0.69	-0.71	0	-7.830
58-40-2	Promazine	-0.55	-0.57	0	-7.530
84625-61-6	Itraconazole	-0.65	-0.67	0	-7.750
146-54-3	Triflupromazine	-0.68	-0.71	0	-7.820
50-53-3	Chlorpromazine	-0.62	-0.64	0	-7.690
91161-71-6	Terbinafine	-1.03	-1.07	0	-8.590
23593-75-1	Clotrimazole	-1.32	-1.36	0	-9.220
131-56-6	2,4-dihydroxybenzophenone; bp-1	-1.44	-1.48	0	-9.480
131-55-5	2,2',4,4'-tetrahydroxybenzophenone; bp-2	-1.42	-1.46	0	-9.440
4065-45-6	Sulisobenzone; BP-4	-1.76	-1.81	0	-10.190
131-54-4	2,2'-dihydroxy-4,4'-dimethoxybenzophenone; BP-6	-1.30	-1.34	0	-9.180
85-19-8	5-chloro-2-hydroxybenzophenone; BP-7	-1.34	-1.38	0	-9.260
131-53-3	Dioxybenzone; BP-8	-1.34	-1.38	0	-9.270
1641-17-4	Mexenone; BP-10	-1.32	-1.36	0	-9.230
611-99-4	4,4'-dihydroxybenzophenone; dhbp	-1.41	-1.46	0	-9.430
117-99-7	2-hydroxybenzophenone; 2-hbp	-1.36	-1.40	0	-9.310
13020-57-0	3-hydroxybenzophenone; 3-hbp	-1.38	-1.42	0	-9.350
1137-42-4	4-hydroxybenzophenone; 4-hbp	-1.44	-1.48	0	-9.480
1143-72-2	2,3,4-trihydroxybenzophenone; 2,3,4-thbp	-1.25	-1.29	0	-9.080
120-47-8	Ethyl 4-hydroxybenzoate; Ethylparaben	-1.48	-1.53	0	-9.580
94-13-3	Propyl 4-hydroxybenzoate; Propylparaben	-1.48	-1.53	0	-9.580
94-26-8	Butyl 4-hydroxybenzoate; Butylparaben	-1.48	-1.53	0	-9.580
140-66-9	4-tert-octylphenol; 4-tert-OP	-1.09	-1.13	0	-8.720
57-63-6	17α-ethinylestradiol; EE2	-1.09	-1.12	0	-8.710
56-53-1	Diethylstilbestrol; DES	-1.12	-1.15	0	-8.780
58-18-4	17α-methyltestosterone; MT	-1.55	-1.60	0	-9.730
13311-84-7	Flutamide	-1.66	-1.71	0	-9.970
134-62-3	N, N-diethyl-m-toluamide; DEET	-1.37	-1.41	0	-9.330
95-14-7	1H-benzotriazole; BT	-1.42	-1.47	0	-9.450
136-85-6	5-methyl-1H-benzotriazole; 5-Me-BT	-1.33	-1.37	0	-9.250
61-68-7	Mefenamic acid	-0.97	-1.00	0	-8.450
42399-41-7	Diltiazem	-1.06	-1.10	0	-8.660
83881-51-0	Cetirizine	-2.39	-2.47	1	-9.170
58-73-1	Diphenhydramine	-1.14	-1.17	0	-8.820
68-88-2	Hydroxyzine	-1.20	-1.23	0	-8.950
6138-79-0	Triprolidine	-1.11	-1.14	0	-8.760
79-06-1	Acrylamide	-1.87	-1.93	0	-10.450
2921-88-2	Chlorpyrifos	-1.42	-1.46	0	-9.440
62-73-7	Dichlorvos	-1.43	-1.47	0	-9.460
50-99-7	Glucose	-1.98	-2.05	0	-10.690
52645-53-1	Permethrin	-1.19	-1.23	0	-8.940
3332-27-2 *	**N,N-dimethyltetradecylamine N-oxide**	**-1.23**	**-1.27**	**0**	**-9.020**
77-58-7	Dibutyltin dilaurate	-1.74	-1.79	0	-10.150
116-37-0	1,1'-isopropylidenebis(p-phenyleneoxy)dipropan-2-ol	-1.08	-1.12	0	-8.700
10222-01-2	2,2-dibromo-2-cyanoacetamide	-2.20	-2.27	0	-11.160
6021-61-0	2-[4-[(2-chloro-4-nitrophenyl)azo]-n-(2-cyanoethyl)anilino]ethyl acetate	-1.14	-1.18	0	-8.830
50-29-3	Dichlorodiphenyltrichloroethane (DDT)	-1.50	-1.55	0	-9.630
309-00-2	Aldrin	-1.58	-1.64	0	-9.810
36355-01-8	Hexabromobiphenyl	-1.58	-1.64	0	-9.810
101-14-4	2,2'-dichloro-4,4'-methylendianiline	-0.96	-0.99	0	-8.420
31508-00-6	1,2,4-trichloro-5-(3,4-dichlorophenyl)benzene	-1.51	-1.56	0	-9.640
208-96-8	Acenaphthylene	-1.26	-1.30	0	-9.100
56-55-3	Benzo[a]anthracene	-0.96	-0.99	0	-8.420
53-70-3	Dibenzo[a,h]anthracene	-0.97	-1.00	0	-8.450
101-55-3	1-bromo-4-phenoxybenzene	-1.25	-1.29	0	-9.070

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
24017-47-8	Triazophos	-1.31	-1.35	0	-9.200
1461-25-2	Tetra-n-butyltin	-1.44	-1.49	0	-9.490
4640-01-1	Methyl triclosan	-1.35	-1.39	0	-9.290
112-18-5	N,N-dimethyldodecan-1-amine	-1.02	-1.06	0	-8.570
124-19-6	Nonanal	-1.71	-1.76	0	-10.080
526-73-8	1,2,3-trimethylbenzene	-1.20	-1.24	0	-8.960
629-59-4	Tetradecane	-2.02	-2.08	0	-10.760
1120-21-4	Undecane	-2.05	-2.12	0	-10.840
84-65-1	9,10-anthracenedione	-1.78	-1.84	0	-10.240
86-74-8	Carbazole	-0.92	-0.95	0	-8.340
92-06-8	1,3-diphenylbenzene	-1.24	-1.28	0	-9.050
580-51-8	3-phenylphenol	-1.24	-1.28	0	-9.040
110-54-3	N-hexane	-2.12	-2.19	0	-11.000
122-88-3	4-chlorophenoxyacetic acid	-2.40	-2.47	1	-9.180
20859-73-8	Aluminum phosphide	-0.53	-0.55	0	-7.480
80060-09-9	Diafenthiuron	-0.92	-0.95	0	-8.340
59-30-3	Folic acid	-3.33	-3.43	2	-8.810
142469-14-5	Tritosulfuron	-2.09	-2.15	0	-10.920
84030-86-4	Esbiothrin	-1.50	-1.55	0	-9.620
5836 10 2	Chlorpropylate	-1.48	-1.53	0	-9.580
78-34-2	Dioxathion	-1.37	-1.42	0	-9.340
957-51-7	Diphenamid	-1.33	-1.37	0	-9.250
2540-82-1	Formothion	-1.54	-1.58	0	-9.700
961-22-8	Azinphosmethyl oxon	-1.58	-1.64	0	-9.810
16655-82-6	3-hydroxycarbofuran	-1.18	-1.21	0	-8.910
3739-38-6	3-Phenoxybenzoic acid	-1.31	-1.36	0	-9.210
107-49-3	Tetraethyl pyrophosphate	-2.19	-2.26	0	-11.150
7421-93-4	Endrin aldehyde	-1.88	-1.94	0	-10.460
31972-43-7	Fenamiphos sulfoxide	-0.88	-0.91	0	-8.260
31972-44-8	Fenamiphos sulfone	-1.54	-1.59	0	-9.710
2581-34-2	3-methyl-4-nitrophenol	-1.70	-1.75	0	-10.060
3761-41-9	Fenthion sulfoxide	-0.94	-0.97	0	-8.390
3761-42-0	Fenthion sulfone	-1.58	-1.63	0	-9.800
87237-48-7	Haloxyfop-2-ethoxyethyl	-1.36	-1.41	0	-9.320
2635 10 1	Methiocarb sulfoxide	-0.86	-0.89	0	-8.220
2179-25-1	Methiocarb sulfone	-1.56	-1.61	0	-9.750
2588 03 6	Phorate sulfoxide	-1.00	-1.03	0	-8.510
2588 04 7	Phorate sulfone	-1.43	-1.48	0	-9.470
1942-71-8	2-(4-tert-butylphenoxy)cyclohexanol	-1.07	-1.10	0	-8.670
27304-13-8	Oxychlordane	-1.68	-1.73	0	-10.020
53380-22-6	Ethiofencarb sulfoxide	-0.90	-0.93	0	-8.290
53380-23-7	Ethiofencarb sulfone	-1.54	-1.59	0	-9.720
311-45-5	Ethyl paraoxon	-1.71	-1.77	0	-10.090
40020-01-7	Pyridafol	-1.32	-1.36	0	-9.230
2703-37-9	Thiometon sulfoxide	-1.10	-1.13	0	-8.730
20301-63-7	Thioometon sulfone	-1.47	-1.51	0	-9.550
95-69-2	4-chloro-2-methylaniline	-0.98	-1.01	0	-8.470
140-38-5	(4-chlorophenyl)urea	-1.18	-1.21	0	-8.910
61898-95-1	Methyl-3-(2,2-dichlorovinyl)-2,2-dimethyl-(1-cyclopropane) carboxylate	-1.61	-1.66	0	-9.870
1713-15-1	2,4-D-1-isobutyl ester	-1.49	-1.53	0	-9.590
62610-77-9	Methacrifos	-1.59	-1.64	0	-9.820
100760-10-9	Quizalafop ethyl	-1.26	-1.30	0	-9.100
2227-13-6	Tetrasul	-1.31	-1.36	0	-9.210
950-10-7	Mephospholan	-1.30	-1.34	0	-9.180
1214-39-7	6-benzyladenine	-1.20	-1.24	0	-8.960
120923-37-7	Amidosulfuron	-1.69	-1.74	0	-10.040

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
120162-55-2	Azimsulfuron	-1.60	-1.65	0	-9.840
120-23-0	2-Naphthyloxyacetic acid	-2.20	-2.27	1	-8.740
41483-43-6	Bupirimate	-1.07	-1.11	0	-8.680
55285-14-8	Carbosulfan	-0.90	-0.93	0	-8.300
1134-23-2	Cycloate	-1.24	-1.28	0	-9.040
13684-56-5	Desmedipham	-1.24	-1.28	0	-9.050
3347-22-6	Dithianon	-1.24	-1.28	0	-9.050
126801-58-9	Ethoxysulfuron	-1.26	-1.30	0	-9.090
61213-25-0	Fluorochloridone	-1.40	-1.45	0	-9.410
77-06-5	Gibberellic acid	-2.91	-3.01	1	-10.320
10004-44-1	Hymexazol	-1.54	-1.59	0	-9.710
81405-85-8	Imazamethabenz-methyl	-1.54	-1.59	0	-9.720
140923-17-7	Iprovalicarb	-1.29	-1.33	0	-9.150
7704-34-9	Sulfur	-1.42	-1.46	0	-9.440
123-33-1	Maleic hydrazide	-1.44	-1.49	0	-9.490
133408-50-1	Metominostrobin	-1.26	-1.30	0	-9.100
2310-17-0	Phosalone	-1.37	-1.41	0	-9.330
90717-03-6	Quinmerac	-1.49	-1.53	0	-9.590
111872-58-3	Halfenprox	-1.20	-1.23	0	-8.950
90035-08-8	Flocoumafen	-1.21	-1.25	0	-8.990
560121-52-0	Cyenopyrafen	-1.22	-1.26	0	-9.000
65731-84-2	Beta cypermethrin	-1.33	-1.37	0	-9.240
56073-10-0	Brodifacoum	-1.29	-1.34	0	-9.170
1469-48-3	Cis-1,2,3,6-tetrahydrophthalimide	-1.71	-1.76	0	-10.080
6515-38-4	3,5,6-trichloro-2-pyridinol	-1.39	-1.43	0	-9.370
1031-07-8	Endosulfan sulfate	-1.80	-1.86	0	-10.280
120068-36-2	Fipronil sulfone	-1.78	-1.83	0	-10.230
120067-83-6	Fipronil sulfide	-1.64	-1.70	0	-9.940
1689-83-4	Ioxynil	-1.54	-1.58	0	-9.700
1646-87-3	Aldicarb-sulfoxide	-1.00	-1.03	0	-8.510
1646-88-4	Aldicarb-sulfone	-1.77	-1.82	0	-10.210
3032-40-4	Fluometuron desmethyl	-1.29	-1.34	0	-9.170
1570-64-5	2-methyl-4-chlorophenol	-1.20	-1.23	0	-8.950
94-80-4	2,4-D-1 -butyl ester	-1.49	-1.54	0	-9.600
789-02-6	o,p'-DDT	-1.49	-1.53	0	-9.590
67564-91-4	Fenpropimorph	-1.06	-1.10	0	-8.660
319-84-6	HCH-alpha	-2.17	-2.24	0	-11.100
319-85-7	HCH-delta	-2.21	-2.28	0	-11.190
103055-07-8	Lufenuron	-1.38	-1.42	0	-9.350
119168-77-3	Tebufenpyrad	-1.36	-1.41	0	-9.320
16484-77-8	Mecoprop-P	-2.51	-2.59	1	-9.420
1746-81-2	Monolinuron	-1.16	-1.20	0	-8.880
52888-80-9	Prosulfocarb	-1.22	-1.26	0	-9.000
52315-07-8	Zeta-cypermethrin	-1.31	-1.35	0	-9.200
66841-25-6	Tralomethrin	-1.38	-1.42	0	-9.350
563-12-2	Ethion	-1.38	-1.42	0	-9.350
70124-77-5	Flucythrinate	-1.39	-1.44	0	-9.390
52918-63-5	Deltamethrin	-1.32	-1.36	0	-9.230
139968-49-3	Metaflumizone	-1.36	-1.40	0	-9.310
70630-17-0	Metalaxyl-M	-1.41	-1.45	0	-9.420
108-62-3	Metaldehyde	-1.99	-2.06	0	-10.710
422556-08-9	Pyroxsulam	-1.32	-1.36	0	-9.220
87820-88-0	Tralkoxydim	-1.20	-1.24	0	-8.960
43121-43-3	Triadimefon	-1.36	-1.41	0	-9.320
55219-65-3	Triadimenol	-1.36	-1.41	0	-9.320
2303-17-5	Triallate	-1.38	-1.43	0	-9.360

Table D.1. Continued.

CAS	Name	Pred $pEC_{50, AB}$	Cal pLC_{50}	nRCOOH	E_{HOMO}
52-68-6	Trichlorphon (Chlorphos)	-1.86	-1.92	0	-10.420
80844-07-1	Etofenprox	-0.94	-0.97	0	-8.390
102851-06-9	Tau-fluvalinate	-1.22	-1.26	0	-9.000

*From this compound to the end: Chemicals with no ecotoxicological data (SU0303, 2015).

APPENDIX E: DETAILED RESULTS OF *Dugesia japonica* MODELS

Table E.1. *D. japonica* external set chemicals, predicted pLC_{50}, model descriptors, production volume status.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
56	110-82-7	Cyclohexane*, §	-2.95	-2.05	3.44	0.000	0.961	0	1.064	-1.66	HPV
57	108-87-2	Methylcyclohexane*, §	-2.81	-1.63	3.61	0.000	0.964	0	1.559	-1.18	HPV
58	1678-91-7	Ethylcyclohexane*	-2.58		4.56	0.000	0.961	0	1.493		HPV
59	109-69-3	1-chlorobutane*, §	-3.16	-3.47	2.64	0.000	0.967	0	0.889	-3.31	HPV
60	75-09-2	Dichloromethane*	-3.17		1.25	0.000	1.040	0	0.177		HPV
61	75-34-3	1,1-dichloroethane*	-3.20		1.79	0.000	1.017	0	-0.017		HPV
62	107-06-2	1,2-dichloroethane*	-3.26		1.48	0.000	1.019	0	0.060		HPV
63	78-87-5	1,2-dichloropropane*, §	-3.22	-2.89	1.98	0.000	0.999	0	0.281	-2.64	HPV
64	142-28-9	1,3-dichloropropane*, §	-3.26	-3.54	2.00	0.000	0.993	0	0.285	-3.39	N
65	67-66-3	Trichloromethane*	-2.29		1.97	0.000	1.149	0	-0.078		HPV
66	79-00-5	1,1,2-trichloroethane*	-2.72		1.89	0.000	1.081	0	0.204		HPV
67	96-18-4	1,2,3-trichloropropane*, §	-2.85	-2.65	2.27	0.000	1.041	0	0.365	-2.36	HPV
68	56-23-5	Carbon tetrachloride*	-1.46		2.83	0.000	1.247	0	-0.461		HPV
69	79-34-5	1,1,2,2-tetrachloroethane*	-2.28		2.39	0.000	1.135	0	-0.182		HPV
70	76-01-7	Pentachloroethane*	-1.68		3.22	0.000	1.188	0	-0.077		HPV
71	109-64-8	1,3-dibromopropane*	-2.95		2.37	0.000	1.018	0	0.553		N
72	75-27-4	Bromodichloromethane*	-2.00		2.00	0.000	1.188	0	0.062		N
73	124-48-1	Dibromochloromethane*	-1.71		2.16	0.000	1.221	0	0.235		N
74	96-12-8	1,2-dibromo-3-chloropropane*	-2.54		2.96	0.000	1.060	0	0.339		N
75	542-75-6	1,3-dichloropropene*, §	-3.02	-1.90	2.04	0.000	1.019	0	0.673	-1.49	HPV
76	760-23-6	3,4-dichlorobut-1-ene*, §	-2.94	-2.07	2.60**	0.000	1.009	0	0.602	-1.69	HPV
77	79-01-6	Trichloroethylene*	-2.34		2.42	0.000	1.113	0	0.320		HPV
78	127-18-4	Tetrachloroethene*	-1.61		3.40	0.000	1.184	0	0.277		HPV
79	78-79-5	Isoprene*	-3.33		2.42	0.000	0.944	0	1.224		HPV
80	111-78-4	1,5-cyclooctadiene*	-3.01		3.16	0.000	0.956	0	1.400		HPV
81	3048-65-5	3a,4,7,7a-tetrahydro-1H-indene*	-2.69		3.28**	0.000	1.008	0	1.035		HPV
82	16219-75-3	5-ethylidene-8,9,10-trinorborn-2-ene*, §	-2.62	-2.80	3.82	0.000	0.997	0	1.012	-2.54	HPV
83	71-36-3	1-butanol*	-4.14		0.88	0.000	0.906	0	0.299		HPV
84	78-83-1	Iso-butanol*	-4.03		0.76	0.000	0.925	0	0.410		HPV
85	75-65-0	2-methyl-2-propanol*	-4.04		0.35	0.000	0.929	0	0.997		HPV
86	71-41-0	1-pentanol*	-3.84		1.51	0.000	0.918	0	0.638		HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
87	584-02-1	3-pentanol*, §	-3.85	-3.35	1.21	0.000	0.932	0	0.503	-3.17	N
88	111-27-3	Hexanol*, §	-3.37	-2.86	2.03	0.205	0.927	0	0.860	-2.60	HPV
89	111-70-6	1-heptanol*, §	-2.45	-2.97	2.62	0.760	0.940	0	1.123	-2.73	N
90	111-87-5	1-octanol*, §	-2.05	-2.56	3.00	0.840	0.951	1	1.247	-2.26	HPV
91	143-08-8	1-nonanol*	-1.67		3.77	0.882	0.963	1	1.512		HPV
92	112-30-1	1-decanol*, §	-1.31	-1.53	4.57	0.907	0.973	1	1.806	-1.06	HPV
93	25339-17-7	Isodecyl alcohol*	-2.39		3.71**	0.907	0.859	1	1.257		HPV
94	108-93-0	Cyclohexanol*	-3.70		1.23	0.000	0.961	0	0.154		HPV
95	96-23-1	1,3-dichloro-2-propanol*	-3.70		0.78**	0.000	0.987	0	-0.123		N
96	107-21-1	Ethylene glycol*	-4.94		-1.36	0.000	0.904	0	-0.783		HPV
97	80-04-6	Hydrogenatedbisphenol A*	-1.41		4.55**	0.815	0.944	4	0.952		N
98	109-86-4	2-methoxyethanol*	-4.86		-0.77	0.000	0.893	0	-0.851		HPV
99	110-80-5	2-ethoxyethanol*	-4.83		-0.32	0.000	0.877	0	-0.778		HPV
100	109-59-1	2-isopropoxyethanol*	-4.66		0.05	0.000	0.879	0	-0.405		N
101	111-76-2	2-butoxyethanol*	-3.36		0.83	0.799	0.895	0	0.037		HPV
102	111-90-0	2-(2 ethoxyethoxy)ethanol*	-3.38		-0.54	1.396	0.843	1	-0.497		HPV
103	112-34-5	2-(2-butoxyethoxy)ethanol*	-3.39		0.56	1.007	0.855	1	-0.204		HPV
104	60-29-7	Diethylether*	-4.22		0.89	0.000	0.895	0	0.273		HPV
105	142-96-1	1,1'-oxybis-butane*, §	-2.09	-2.60	3.21	0.907	0.932	0	1.668	-2.30	HPV
106	111-44-4	Bis(2-chloroethyl) ether*, §	-3.77	-3.39	1.29	0.000	0.949	0	0.089	-3.22	HPV
107	127-90-2	2,3,3,3,2',3',3',3'-octachlorodipropyl ether*	-0.95		5.10**	0.947	1.040	0	0.702		N
108	75-07-0	Acetaldehyde*	-4.54		-0.34	0.000	0.903	0	0.057		HPV
109	123-15-9	2-methylvaleraldehyde*	-3.46		1.73**	0.000	0.952	0	1.237		N
110	4170-30-3	Crotonaldehyde*	-3.89		0.60**	0.000	0.957	0	0.253		HPV
111	111-30-8	Glutaraldehyde*	-4.39		-0.33	0.000	0.937	0	-0.490		HPV
112	67-64-1	Acetone*	-4.27		-0.24	0.000	0.929	0	0.524		HPV
113	693-54-9	2-decanone*	-1.41		3.73	0.894	1.005	0	2.167		N
114	112-12-9	2-undecanone*, §	-1.13	-0.73	4.09	0.912	1.012	0	2.902	-0.13	N
115	593-08-8	2-tridecanone*	-0.85		4.68**	0.934	1.025	0	2.943		N
116	108-94-1	Cyclohexanone*	-3.59		0.81	0.000	0.976	0	1.025		HPV
117	1502-22-3	2-(1'-cyclohexenyl)cyclohexanone*	-2.12		3.17	0.734	0.954	0	1.924		N
118	78-59-1	3,5,5-trimethyl-2-cyclohexen-1-one*, §	-3.30	-3.15	1.70	0.000	0.971	0	1.555	-2.94	HPV
119	141-78-6	Ethylacetate*	-4.11		0.73	0.000	0.930	0	-0.297		HPV
120	110-19-0	Isobutyl acetate*	-3.69		1.78	0.000	0.932	0	0.506		HPV
121	111-82-0	Methyl dodecanoate*	-0.77		5.41	0.940	1.028	0	1.894		HPV
122	515-84-4	Ethyl trichloroacetate*	-3.20		2.39	0.000	1.002	0	-0.491		N
123	105-53-3	Diethyl malonate*, §	-3.50	-3.29	0.96	0.551	0.937	0	-1.132	-3.10	HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
124	96-33-3	Methyl acrylate*, §	-3.94	-1.89	0.80	0.000	0.957	0	-0.498	-1.48	HPV
125	140-88-5	Ethylacrylate*, §	-3.80	-2.16	1.32	0.000	0.946	0	0.023	-1.79	HPV
126	141-32-2	n-butyl acrylate*, §	-2.22	-2.18	2.36	0.986	0.960	0	0.542	-1.81	HPV
127	818-61-1	2-hydroxyethyl acrylate*, §	-4.18	-0.86	-0.21	0.297	0.934	0	-1.417	-0.28	HPV
128	97-88-1	n-butyl methacrylate*	-2.10		2.88	0.986	0.950	0	0.808		HPV
129	688-84-6	2-ethylhexyl methacrylate*	-1.59		4.54	1.028	0.925	0	1.927		HPV
130	868-77-9	2-hydroxyethyl methacrylate*	-4.02		0.47	0.280	0.928	0	-1.246		HPV
131	2867-47-2	2-(dimethylamino)ethyl methacrylate*, §	-3.53	-2.45	0.97**	0.520	0.897	0	0.725	-2.13	HPV
132	13048-33-4	Hexamethylene diacrylate*	-2.80		3.08**	0.851	0.856	0	0.926		HPV
133	108-05-4	Vinyl acetate*	-3.99		0.73	0.000	0.942	0	-0.032		HPV
134	600-07-7	2-methylbutanoic acid*	-3.81		1.49**	0.000	0.944	0	-0.284		HPV
135	503-74-2	3-methylbutanoic acid*	-3.79		1.16	0.000	0.948	0	0.305		N
136	75-98-9	Pivalic acid*	-3.81		1.48	0.000	0.944	0	-0.283		HPV
137	88-09-5	2-ethyl-butanoic acid*	-3.79		1.68	0.000	0.947	0	-0.639		N
138	111-14-8	Heptanoic acid*	-2.31		2.42	0.722	0.986	0	0.654		HPV
139	124-07-2	Octanoic acid*	-1.82		3.05	0.821	0.994	1	0.925		HPV
140	334-48-5	Decanoic acid*	-2.36		4.09	0.898	0.854	1	0.995		HPV
141	79-11-8	Chloroacetic acid*, §	-3.94	-3.12	0.22	0.000	0.988	0	-0.727	-2.91	HPV
142	335-67-1	Perfluorooctanoic acid*	-0.74		4.81**	0.809	1.050	0	3.300		N
143	298-12-4	Glyoxylic acid*	-4.68		-1.40**	0.000	0.952	0	-1.118		HPV
144	3821-81-6	A-fluoro-β-alanine*	-5.51		-4.23**	0.000	0.945	0	-1.070		N
145	79-10-7	Acrylic acid*, §	-4.08	-3.04	0.35	0.000	0.950	0	-0.277	-2.81	HPV
146	79-41-4	Methacrylic acid*, §	-3.98	-3.41	0.93	0.000	0.954	0	-0.885	-3.24	HPV
147	110-44-1	Sorbic acid*	-3.24		1.33	0.283	0.996	0	-0.616		HPV
148	144-62-7	Oxalic acid*, §	-4.98	-3.36	-1.74**	0.000	0.940	0	-1.952	-3.18	HPV
149	124-04-9	Adipic acid*, §	-3.81	-3.00	0.08	0.480	0.921	0	-0.315	-2.77	HPV
150	108-91-8	Cyclohexylamine*	-3.52		1.49	0.000	0.961	0	0.876		HPV
151	141-43-5	Monoethanolamine*	-4.87		-1.31	0.000	0.900	0	-0.256		HPV
152	115-70-8	2-amino-2-ethylpropanediol*	-4.60		-0.6**	0.000	0.917	0	-0.509		N
153	109-89-7	Diethylamine*, §	-4.22	-3.10	0.58	0.000	0.895	0	0.814	-2.88	HPV
154	111-42-2	Diethanolamine/2,2'-iminodiethanol*, §	-5.05	-3.19	-1.43	0.013	0.873	0	-0.122	-2.99	HPV
155	121-44-8	Triethylamine*	-3.88		1.45	0.000	0.903	0	1.162		HPV
156	102-81-8	2-(dibutylamino)ethanol*	-2.58		2.01**	0.858	0.917	0	1.688		HPV
157	124-09-4	1,6-hexanediamine*	-3.07		0.35**	0.577	0.970	0	1.281		HPV
158	6864-37-5	2,2'-dimethyl-4,4'-methylenebis(cyclohexylamine)*, §	-1.02	-2.10	4.10**	1.122	0.936	4	2.451	-1.72	HPV
159	111-18-2	N,N,N',N'-tetramethylhexamethylenediamine*	-2.30		1.70**	0.895	0.950	0	2.402		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
160	3030-47-5	N-methyl-N,N-bis(2-dimethylaminoethyl)amine*	-3.29		-0.57**	0.668	0.944	0	1.971		HPV
161	629-40-3	Octanedinitrile*	-2.74		0.59	0.781	0.977	0	1.211		N
162	107-13-1	2-propenenitrile*, §	-4.06	-2.53	0.25	0.000	0.945	0	0.270	-2.22	HPV
163	126-98-7	Methacrylonitrile*	-3.83		0.68	0.000	0.960	0	0.343		N
164	920-37-6	2-propenenitrile, 2-chloro-*	-3.51		0.70**	0.000	1.016	0	0.027		N
165	406-86-0	4,4,4-trifluorocrotonitrile*	-2.71		1.63**	0.000	1.015	0	3.709		N
166	1855-63-6	1-cyclohexene-1-carbonitrile*	-3.20		2.04**	0.000	0.997	0	0.397		N
167	1118-61-2	3-amino-2-Butenenitrile*	-4.42		-0.79**	0.000	0.950	0	-0.428		N
168	764-42-1	2-butenedinitrile, (e)-*	-3.84		-0.25	0.000	1.004	0	0.104		N
169	75-91-2	Tert-butylhydroperoxide*	-3.97		0.94**	0.000	0.927	0	0.435		HPV
170	3006-82-4	Tert-butyl 2-ethylperoxyhexanoate*	-1.84		4.31**	1.256	0.876	0	1.096		HPV
171	76-06-2	Trichloronitromethane*	-2.22		2.09	0.000	1.135	0	0.804		N
172	96-29-7	2-butanone oxime*, §	-4.03	-3.25	0.63	0.000	0.934	0	0.263	-3.06	HPV
173	100-64-1	Cyclohexanone oxime*	-3.69		0.84	0.000	0.971	0	0.544		HPV
174	60-34-4	Methylhydrazine*	-4.86		-1.05	0.000	0.887	0	-0.036		N
175	57-14-7	N,N-dimethylhydrazine*	-4.71		-1.19**	0.000	0.906	0	0.350		N
176	657-24-9	Metformin*, §	-5.11	-1.47	-2.64**	0.000	0.923	0	-0.366	-0.99	N
177	110-91-8	Morpholine*	-4.31		-0.86	0.000	0.957	0	0.154		HPV
178	2403-88-5	2,2,6,6-tetramethylpiperidin-4-ol*, §	-3.83	-1.47	0.94**	0.000	0.946	0	0.490	-0.99	HPV
179	110-85-0	Piperazine*, §	-4.38	-2.68	-1.50	0.000	0.962	0	0.654	-2.39	HPV
180	108-80-5	Isocyanuric acid*	-3.41		0.61**	0.000	1.022	0	0.577		HPV
181	470-82-6	2-oxabicyclo[2.2.2]octane, 1,3,3-trimethyl-*	-3.11		2.74	0.000	0.960	0	1.345		N
182	15045-43-9	2,2,5,5-tetramethyltetrahydrofuran*	-3.30		2.06	0.000	0.953	0	1.645		N
183	62571-86-2	Captopril*	-2.97		0.34	1.062	0.950	0	-0.890		N
184	674-82-8	But-3-en-3-olide*	-3.87		-0.39**	0.000	1.002	0	0.236		HPV
185	106-91-2	Glycidyl methacrylate*, §	-2.77	-2.55	0.81**	1.046	0.945	0	-0.072	-2.24	HPV
186	75-08-1	Ethanethiol*	-3.67		1.27**	0.000	0.958	0	0.418		HPV
187	110-66-7	Pentane-1-thiol*	-3.04		2.74**	0.000	0.976	0	1.144		N
188	111-88-6	1-mercaptooctane <n-octylmercaptan>*	-1.24		4.21**	0.998	0.997	0	1.978		HPV
189	143-10-2	1-decanethiol*	-0.86		5.2**	0.935	1.011	0	2.495		N
190	60-24-2	2-mercaptoethanol*	-4.30		-0.20**	0.000	0.943	0	-0.372		HPV
191	624-92-0	Dimethyl disulphide*	-3.05		1.77	0.000	1.034	0	0.259		HPV
192	110-81-6	Diethyl disulfide*	-3.24		2.86**	0.000	0.948	0	0.785		N
193	3268-49-3	3-(methylthio)propionaldehyde*, §	-4.06	-1.77	0.41**	0.000	0.945	0	0.024	-1.34	HPV
194	111-17-1	3,3'-thiodipropionic acid*	-3.99		-0.18**	0.427	0.948	0	-1.839		N
195	556-61-6	Methyl isothiocyanate*	-3.65		0.94	0.000	0.987	0	-0.087		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
196	79-19-6	Thiosemicarbazide*	-4.79		-1.67**	0.000	0.944	0	-0.986		N
197	4189-44-0	Thiourea dioxide*	-4.45		-2.64**	0.000	1.013	0	0.094		HPV
198	1763-23-1	Perfluorooctane sulfonic acid*	-0.91		4.49**	0.935	0.979	0	4.994		N
199	115-96-8	Tris(2-chloroethyl) phosphate*	-3.10		1.44	0.674	0.924	0	0.313		HPV
200	71-43-2	Benzene*, §	-3.18	-3.43	2.13	0.000	1.000	0	0.251	-3.27	HPV
201	108-88-3	Toluene*, §	-2.99	-2.62	2.73	0.000	0.998	0	0.496	-2.33	HPV
202	100-41-4	Ethylbenzene*, §	-2.96	-1.74	3.15	0.000	0.982	0	0.643	-1.30	HPV
203	95-47-6	o-xylene*, §	-2.93	-2.44	3.12	0.000	0.984	0	0.788	-2.12	HPV
204	108-38-3	m-xylene*, §	-2.83	-2.64	3.20	0.000	0.998	0	0.697	-2.35	HPV
205	106-42-3	p-xylene*, §	-2.83	-2.70	3.15	0.000	0.998	0	0.763	-2.42	HPV
206	103-65-1	n-propylbenzene*, §	-2.01	-1.70	3.69	0.648	0.984	0	0.983	-1.26	HPV
207	98-82-8	Isopropylbenzene*, §	-2.81	-1.22	3.66	0.000	0.973	0	1.050	-0.70	HPV
208	104-51-8	Butylbenzene*, §	-1.51	-1.04	4.38	0.904	0.984	0	1.087	-0.49	N
209	99-87-6	p-cymene*, §	-0.85	-2.07	4.10	1.577	0.975	0	1.340	-1.69	HPV
210	98-51-1	4-tert-butyltoluene*	-0.14		5.17	1.994	0.968	0	1.268		HPV
211	25321-09-9	Diisopropylbenzene*, §	-2.81	-0.94	4.10	0.000	0.943	0	1.559	-0.37	HPV
212	827-52-1	Cyclohexylbenzene*	-1.41		4.81**	0.815	0.985	0	1.618		N
213	108-90-7	Chlorobenzene*, §	-2.77	-2.24	2.84	0.000	1.035	0	0.136	-1.88	HPV
214	95-49-8	2-chlorotoluene*, §	-2.64	-2.51	3.42	0.000	1.022	0	0.478	-2.20	HPV
215	108-41-8	3-chlorotoluene*	-2.70		3.28	0.000	1.024	0	0.254		N
216	106-43-4	4-chlorotoluene*, §	-2.68	-1.87	3.33	0.000	1.023	0	0.322	-1.45	HPV
217	95-50-1	1,2-dichlorobenzene*, §	-2.40	-1.63	3.43	0.000	1.068	0	0.020	-1.18	HPV
218	541-73-1	1,3-dichlorobenzene*, §	-2.41	-1.53	3.53	0.000	1.069	0	-0.243	-1.06	HPV
219	106-46-7	1,4-dichlorobenzene*, §	-2.39	-1.63	3.44	0.000	1.069	0	0.030	-1.18	HPV
220	95-73-8	2,4-dichlorotoluene*, §	-2.31	-1.11	4.24	0.000	1.046	0	0.074	-0.57	HPV
221	95-75-0	3,4-dichlorotoluene*	-2.37		3.95	0.000	1.048	0	0.117		N
222	19398-61-9	2,5-dichlorotoluene*	-2.35		3.97	0.000	1.047	0	0.312		N
223	118-69-4	2,6-dichlorotoluene*	-2.35		3.99	0.000	1.046	0	0.281		N
224	87-61-6	1,2,3-trichlorobenzene*, §	-2.08	-1.57	4.05	0.000	1.099	0	-0.378	-1.11	N
225	120-82-1	1,2,4-trichlorobenzene*, §	-2.08	-1.45	4.02	0.000	1.100	0	-0.354	-0.97	HPV
226	108-86-1	Bromobenzene*, §	-2.65	-1.97	2.99	0.000	1.047	0	0.104	-1.57	N
227	348-61-8	4-bromo-1,2-difluorobenzene*	-2.23		3.28**	0.000	1.071	0	1.340		N
228	98-87-3	Alpha,alpha-dichlorotoluene*	-2.66		2.97**	0.000	1.043	0	0.261		HPV
229	611-19-8	2-Chlorobenzyl chloride*, §	-2.48	-1.77	3.44**	0.000	1.046	0	0.458	-1.34	HPV
230	98-08-8	Benzotrifluoride ((trifluoromethyl)benzene)*	-2.42		3.01	0.000	1.025	0	2.599		HPV
231	402-31-3	Metaxylene hexafluoride*	-1.78		3.83	0.000	1.039	0	4.826		N
232	98-83-9	2-phenylpropene*, §	-2.78	-1.90	3.48	0.000	0.985	0	1.095	-1.49	HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
233	1321-74-0	Divinylbenzene*	-2.74		3.80**	0.000	0.988	0	0.637		HPV
234	100-51-6	Benzyl alcohol*	-3.61		1.10	0.000	0.990	0	-0.256		HPV
235	2100-42-7	2-chlorohydroquinonedimethylether*	-3.17		2.69	0.022	0.986	0	-0.271		HPV
236	93-15-2	4-allyl-1,2-dimethoxybenzene*	-2.51		3.03**	0.654	0.952	0	0.245		N
237	122-57-6	Benzalacetone*	-2.02		2.07	0.954	1.008	0	0.525		N
238	100-52-7	Benzaldehyde*, §	-3.29	-2.65	1.48	0.000	1.014	0	0.082	-2.36	HPV
239	487-68-3	2,4,6-trimethylbenzaldehyde*	-2.81		3.35**	0.014	0.996	0	0.501		N
240	123-11-5	p-methoxybenzaldehyde*	-1.26		1.76	1.907	0.992	0	-0.302		HPV
241	90-02-8	Salicylaldehyde/2-hydroxybenzaldehyde*, §	-3.18	-1.95	1.81	0.000	1.006	0	0.611	-1.55	N
242	98-86-2	Acetophenone*	-3.30		1.58	0.000	0.999	0	0.510		HPV
243	84-66-2	Diethyl phthalate*, §	-2.52	-2.68	2.42	1.156	0.923	0	-1.249	-2.39	HPV
244	84-69-5	Diisobutyl phthalate*	-2.36		4.11	1.246	0.853	0	-0.927		HPV
245	84-74-2	Dibutyl phthalate*, §	-2.42	-1.57	4.50	1.090	0.860	0	-1.222	-1.10	HPV
246	131-17-9	Diallyl phthalate*	-2.32		3.23	1.058	0.911	0	-0.155		HPV
247	65-85-0	Benzoic acid*	-3.39		1.87	0.000	1.013	0	-1.252		HPV
248	99-04-7	m-toluic acid*, §	-3.19	-2.98	2.37	0.000	1.012	0	-0.788	-2.74	HPV
249	99-94-5	4-methylbenzoic acid*	-2.92		2.27	0.309	1.009	0	-1.042		N
250	69-72-7	Salicylic acid*, §	-3.36	-3.07	2.26	0.000	1.005	0	-1.423	-2.85	HPV
251	99-96-7	4-hydroxybenzoic acid*, §	-3.54	-3.19	1.58	0.029	1.012	0	-1.894	-2.99	HPV
252	19715-19-6	3,5-di-tert-butylsalicylic acid*	-1.20		6.06**	1.150	0.957	0	-0.650		N
253	2840-28-0	3-amino-4-chlorobenzoic acid*	-2.13		1.60**	1.185	1.031	0	-2.096		N
254	25812-30-0	Gemfibrozil*, §	-2.17	-2.00	4.77**	0.833	0.877	2	-0.313	-1.60	N
255	94-74-6	2-methyl-4-chlorophenoxyacetic acid*, §	-2.25	-2.94	3.25	0.839	0.972	1	-1.510	-2.70	HPV
256	140-10-3	Trans-cinnamic acid*	-2.32		2.13	0.762	1.014	0	-0.379		N
257	108-95-2	Phenol*, §	-3.43	-2.25	1.46	0.000	1.000	0	-0.169	-1.89	HPV
258	106-44-5	4-cresol*, §	-3.26	-1.57	1.94	0.000	0.998	0	0.145	-1.11	HPV
259	108-39-4	3-methylphenol*, §	-3.28	-2.55	1.96	0.000	0.998	0	0.000	-2.24	HPV
260	95-48-7	2-methylphenol*, §	-3.26	-2.05	1.95	0.000	0.992	0	0.379	-1.66	HPV
261	90-00-6	2-ethylphenol*	-3.25		2.47	0.000	0.974	0	0.276		HPV
262	123-07-9	4-ethylphenol*, §	-2.78	-2.06	2.58	0.344	0.983	0	0.246	-1.67	HPV
263	620-17-7	3-ethylphenol*	-3.25		2.40	0.000	0.982	0	0.088		HPV
264	526-75-0	2,3-dimethylphenol*	-3.16		2.48	0.000	0.985	0	0.389		HPV
265	576-26-1	2,6-dimethylphenol*, §	-3.15	-1.72	2.36	0.000	0.991	0	0.448	-1.28	HPV
266	95-65-8	3,4-dimethylphenol*	-3.24		2.23	0.000	0.985	0	0.324		HPV
267	95-87-4	2,5-dimethylphenol*, §	-3.11	-2.26	2.33	0.000	0.995	0	0.583	-1.91	HPV
268	105-67-9	2,4-dimethylphenol*, §	-3.12	-1.99	2.30	0.000	0.997	0	0.488	-1.59	HPV
269	108-68-9	3,5-dimethylphenol*	-3.11		2.35	0.000	0.999	0	0.340		HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
270	527-60-6	2,4,6-trimethylphenol*, §	-2.98	-1.81	2.73	0.000	0.994	0	0.742	-1.38	HPV
271	697-82-5	2,3,5-trimethylphenol*	-2.97		3.15**	0.000	0.986	0	0.338		N
272	2416-94-6	2,3,6-trimethylphenol*	-3.12		2.67	0.000	0.984	0	0.322		HPV
273	88-18-6	2-tert-butyl phenol*	-3.14		3.31	0.000	0.957	0	0.218		HPV
274	89-72-5	o-sec-butylphenol*	-2.67		3.27	0.451	0.945	0	0.533		HPV
275	89-83-8	Thymol*	-1.12		3.30	1.622	0.968	0	1.013		HPV
276	99-71-8	p-sec-butylphenol*	-2.40		3.08	0.659	0.955	0	0.684		HPV
277	14938-35-3	4-pentylphenol*	-1.24		4.06	1.104	0.991	1	0.938		N
278	88-60-8	6-tert-butyl-m-cresol*	-0.56		3.97**	2.046	0.962	0	0.550		HPV
279	2219-82-1	6-tert-butyl-o-cresol*	-2.91		3.97**	0.000	0.960	0	0.402		HPV
280	2409-55-4	2-tert-butyl-p-cresol*	-2.88		3.97**	0.000	0.962	0	0.519		HPV
281	1879-09-0	2-(1,1-Dimethylethyl)-4,6-dimethylphenol*, §	-2.67	-1.32	4.52**	0.000	0.965	0	0.750	-0.81	N
282	96-76-4	2,4-di-tert-butylphenol*	-0.78		5.19	1.589	0.948	0	0.896		HPV
283	1806-26-4	4-n-octylphenol*, §	-0.79	-0.31	5.5**	0.932	1.007	1	1.853	0.36	N
284	5510-99-6	2,6-di-sec-butylphenol*	-1.63		4.36	1.043	0.924	0	1.952		N
285	120-95-6	2,4-di-tert-pentylphenol*	-0.81		6.31**	1.423	0.916	0	1.231		HPV
286	95-57-8	2-chlorophenol*, §	-3.02	-1.99	2.15	0.000	1.030	0	-0.066	-1.59	HPV
287	108-43-0	3-chlorophenol*	-2.92		2.50	0.000	1.031	0	-0.077		N
288	106-48-9	4-chlorophenol*, §	-2.98	-1.91	2.39	0.000	1.031	0	-0.288	-1.50	HPV
289	59-50-7	4-chloro-3-methylphenol*, §	-2.84	-1.61	3.10	0.000	1.019	0	-0.115	-1.15	HPV
290	576-24-9	2,3-dichlorophenol*	-2.65		2.84	0.000	1.059	0	-0.128		N
291	120-83-2	2,4-dichlorophenol*, §	-2.63	-1.60	3.06	0.000	1.060	0	-0.452	-1.14	HPV
292	583-78-8	2,5-dichlorophenol*	-2.58		3.06	0.000	1.061	0	-0.211		HPV
293	87-65-0	2,6-dichlorophenol*	-2.73		2.75	0.000	1.060	0	-0.579		N
294	95-77-2	3,4-dichlorophenol*	-2.52		3.33	0.000	1.060	0	-0.263		N
295	591-35-5	3,5-dichlorophenol*	-2.47		3.62	0.000	1.061	0	-0.513		N
296	15950-66-0	2,3,4-trichlorophenol*	-2.25		3.80	0.000	1.088	0	-0.558		N
297	933-78-8	2,3,5-trichlorophenol*	-2.23		3.84	0.000	1.089	0	-0.557		N
298	933-75-5	2,3,6-trichlorophenol*	-2.28		3.77	0.000	1.088	0	-0.669		N
299	95-95-4	2,4,5-trichlorophenol*, §	-2.28	-1.45	3.72	0.000	1.089	0	-0.680	-0.96	N
300	88-06-2	2,4,6-trichlorophenol*, §	-2.36	-1.82	3.69	0.000	1.089	0	-1.166	-1.39	HPV
301	58-90-2	2,3,4,6-tetrachlorophenol*	-2.00		4.45	0.000	1.116	0	-1.326		N
302	87-86-5	Pentachlorophenol*	-1.70		5.12	0.000	1.142	0	-1.734		N
303	106-41-2	4-bromophenol*, §	-2.87	-1.94	2.59	0.000	1.041	0	-0.378	-1.54	N
304	615-58-7	2,4-dibromophenol*	-2.43		3.22	0.000	1.080	0	-0.293		N
305	118-79-6	2,4,6-tribromophenol*, §	-1.92	-1.33	4.13	0.000	1.118	0	-0.304	-0.82	HPV
306	1745-81-9	2-allylphenol*, §	-2.36	-1.40	2.91**	0.587	0.979	0	0.797	-0.91	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
307	4286-23-1	4-(1-methylethenyl)phenol*	-2.42		2.96**	0.579	0.986	0	0.015		N
308	90-05-1	2-methoxyphenol*, §	-3.69	-2.62	1.32	0.000	0.977	0	-0.647	-2.32	HPV
309	25013-16-5	Butylated hydroxyanisole*	-1.62		3.5**	1.289	0.961	0	0.096		N
310	88-75-5	2-nitrophenol*, §	-3.23	-2.35	1.79	0.000	1.009	0	0.152	-2.01	HPV
311	89-64-5	4-chloro-2-nitrophenol*	-2.92		2.46	0.000	1.034	0	-0.168		N
312	88-30-2	3-trifluoromethyl-4-nitrophenol*	-2.32		2.87**	0.029	1.020	0	3.578		N
313	123-31-9	Hydroquinone*	-3.70		0.59	0.000	1.001	0	-0.419		HPV
314	121-79-9	Propyl gallate*	-1.83		1.80	1.265	0.984	3	-1.203		N
315	62-53-3	Aniline*, §	-3.71	-1.15	0.90	0.000	0.996	0	-0.830	-0.62	HPV
316	95-53-4	2-methylaniline*, §	-3.57	-1.21	1.32	0.000	0.987	0	-0.280	-0.69	HPV
317	108-44-1	3-methylaniline*, §	-3.54	-1.33	1.40	0.000	0.997	0	-0.669	-0.83	HPV
318	106-49-0	4-methylaniline*	-3.54		1.39	0.000	0.996	0	-0.587		HPV
319	578-54-1	2-ethylaniline*, §	-3.56	-1.98	1.74	0.000	0.970	0	-0.250	-1.58	N
320	587-02-0	3-ethylaniline*	-3.41		2.11**	0.000	0.983	0	-0.509		N
321	589-16-2	4-ethylaniline*, §	-2.81	-1.71	1.96	0.562	0.982	0	-0.458	-1.27	N
322	87-59-2	2,3-dimethylaniline*	-3.32		2.17**	0.000	0.980	0	0.086		N
323	87-62-7	2,6-dimethylaniline*	-3.39		1.84	0.000	0.984	0	0.094		HPV
324	95-64-7	3,4-dimethylaniline*	-3.45		1.84	0.000	0.986	0	-0.417		N
325	95-68-1	2,4-dimethylaniline*	-3.42		1.68	0.000	0.990	0	-0.059		HPV
326	95-78-3	2,5-dimethylaniline*	-3.37		1.83	0.000	0.990	0	-0.060		N
327	108-69-0	3,5-dimethylaniline*	-3.30		2.17**	0.000	0.998	0	-0.549		N
328	579-66-8	2,6-diethylaniline*	-2.58		3.15**	0.521	0.958	0	0.283		HPV
329	99-88-7	4-isopropylaniline*	-2.36		2.49	0.850	0.973	0	-0.194		HPV
330	88-05-1	2,4,6-trimethylaniline*	-3.10		2.72**	0.000	0.988	0	0.206		N
331	95-51-2	2-chloroaniline*, §	-3.21	-1.32	1.90	0.000	1.024	0	-0.601	-0.81	HPV
332	108-42-9	3-chloroaniline*	-3.26		1.88	0.000	1.025	0	-0.931		N
333	106-47-8	4-chloroaniline*	-3.28		1.83	0.000	1.024	0	-0.925		N
334	95-81-8	2-chloro-5-methylaniline*	-3.14		2.27**	0.000	1.017	0	-0.483		N
335	95-76-1	3,4-dichloroaniline*, §	-2.87	-1.58	2.69	0.000	1.051	0	-1.012	-1.12	HPV
336	95-82-9	2,5-dichloroaniline*, §	-2.80	-1.70	2.75	0.000	1.051	0	-0.617	-1.26	N
337	554-00-7	2,4-dichloroaniline*, §	-2.83	-1.39	2.78	0.000	1.051	0	-0.905	-0.90	N
338	608-27-5	2,3-dichloroaniline*	-2.82		2.82	0.000	1.050	0	-0.873		N
339	608-31-1	2,6-dichloroaniline*	-2.77		2.76	0.000	1.050	0	-0.403		N
340	626-43-7	3,5-dichloroaniline*, §	-2.84	-2.45	2.90	0.000	1.052	0	-1.210	-2.13	N
341	634-67-3	2,3,4-trichloroaniline*, §	-2.55	-1.12	3.33	0.000	1.076	0	-1.169	-0.58	N
342	634-93-5	2,4,6-trichloroaniline*	-2.45		3.52	0.000	1.077	0	-0.873		N
343	636-30-6	2,4,5-trichloroaniline*, §	-2.48	-1.04	3.45	0.000	1.077	0	-0.934	-0.49	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
344	634-91-3	3,4,5-trichloroaniline*	-2.57		3.32	0.000	1.077	0	-1.319		N
345	104-94-9	P-anisidine*, §	-2.66	-0.99	0.95	1.040	0.981	0	-1.160	-0.43	HPV
346	88-74-4	2-nitroaniline*	-3.33		1.85	0.000	1.004	0	-0.402		HPV
347	99-09-2	3-nitroaniline*	-3.54		1.37	0.000	1.011	0	-1.244		N
348	100-01-6	4-nitroaniline*	-3.48		1.39	0.033	1.010	0	-1.120		HPV
349	99-52-5	2-methyl-4-nitroaniline*	-3.29		2.02**	0.032	1.001	0	-0.588		N
350	89-63-4	4-chloro-2-nitroaniline*	-3.00		2.72	0.000	1.027	0	-0.862		HPV
351	96-96-8	2-nitro-p-anisidine*	-2.86		1.94	0.546	0.986	0	-0.780		N
352	103-69-5	N-ethylaniline*, §	-2.61	-1.08	2.16	0.684	0.968	0	0.209	-0.54	HPV
353	121-69-7	N,N-dimethylaniline*, §	-3.30	-2.38	2.31	0.000	0.975	0	0.198	-2.05	HPV
354	91-66-7	N,N-diethylaniline*, §	-2.31	-1.48	3.31	0.694	0.949	0	0.873	-1.00	HPV
355	106-50-3	p-phenylenediamine*	-4.19		-0.30	0.000	0.994	0	-1.768		HPV
356	108-45-2	m-phenylenediamine*, §	-4.18	-2.12	-0.33	0.000	0.995	0	-1.656	-1.74	HPV
357	95-54-5	o-phenylenediamine*	-4.06		0.15	0.000	0.988	0	-1.426		HPV
358	95-70-5	2,5-diaminotoluene*	-4.03		0.16**	0.000	0.988	0	-1.293		N
359	95-80-7	2,4-diaminotoluene*	-4.03		0.14	0.000	0.988	0	-1.200		HPV
360	101-96-2	1,4-benzenediamine, N,N'-bis(1-methylpropyl)-*	-2.05		3.50**	0.938	0.921	2	0.174		HPV
361	85068-29-7	3,5-bis(trifluoromethyl)benzylamine*	-2.05		3.00**	0.016	1.014	0	5.526		N
362	1477-55-0	m-phenylenebis(methylamine)*, §	-3.86	-1.69	0.15**	0.085	0.971	0	0.023	-1.24	HPV
363	29122-68-7	Atenolol*	-3.27		0.16	0.653	0.931	2	-0.073		HPV
364	95-55-6	2-aminophenol*	-3.75		0.62	0.000	0.992	0	-0.439		HPV
365	123-30-8	4-aminophenol*, §	-4.02	-0.91	0.04	0.000	0.997	0	-1.355	-0.34	HPV
366	591-27-5	3-aminophenol*, §	-3.91	-3.47	0.21	0.000	0.996	0	-0.888	-3.31	HPV
367	119-34-6	4-amino-2-nitrophenol*	-3.61		0.96	0.000	1.006	0	-0.704		N
368	98-95-3	Nitrobenzene*, §	-3.23	-2.50	1.85	0.000	1.016	0	-0.296	-2.18	HPV
369	88-72-2	2-nitrotoluene*, §	-3.04	-2.19	2.30	0.000	1.002	0	0.817	-1.83	HPV
370	99-99-0	4-methylnitrobenzene*, §	-3.09	-2.33	2.37	0.030	1.010	0	-0.255	-1.99	HPV
371	88-73-3	2-chloronitrobenzene*, §	-2.87	-2.58	2.24	0.000	1.042	0	0.237	-2.28	HPV
372	121-73-3	3-chloronitrobenzene*, §	-2.95	-2.43	2.46	0.000	1.043	0	-0.720	-2.10	HPV
373	100-00-5	4-chloronitrobenzene*, §	-2.91	-2.06	2.39	0.000	1.043	0	-0.346	-1.68	HPV
374	13290-74-9	4-chloro-3-methylnitrobenzene*	-2.76		3.00**	0.000	1.031	0	0.078		N
375	99-54-7	3,4-dichloronitrobenzene*	-2.62		3.12	0.000	1.069	0	-0.921		HPV
376	89-69-0	1,2,4-trichloro-5-nitrobenzene*	-2.31		3.48	0.000	1.094	0	-0.652		HPV
377	350-30-1	2-chloro-1-fluoro-4-nitrobenzene*	-2.81		2.66**	0.000	1.052	0	-0.583		N
378	100-14-1	Alpha-Chloro-4-nitrotoluene*	-1.80		2.61**	1.007	1.029	0	-0.337		N
379	91-23-6	2-nitroanisole*	-3.40		1.73	0.000	0.989	0	0.006		HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
380	555-03-3	3-nitroanisole*	-3.29		2.16	0.021	0.991	0	-0.291		N
381	606-20-2	2,6-dinitrotoluene*, §	-3.12	-2.29	2.10	0.000	1.007	0	0.398	-1.94	N
382	97-00-7	1-chloro-2,4-dinitrobenzene*, §	-2.96	-1.14	2.17	0.000	1.048	0	-0.470	-0.60	HPV
383	534-52-1	4,6-dinitro-o-cresol*, §	-3.11	-1.66	2.13	0.035	1.015	0	-0.246	-1.21	N
384	88-85-7	2-(1-methylpropyl)-4,6-dinitro-Phenol (Dinoseb)*	-2.45		3.56	0.449	0.968	0	0.444		HPV
385	40487-42-1	Pendimethalin*	-1.39		5.20	1.064	0.943	0	0.971		HPV
386	1582-09-8	Trifluralin*, §	-1.07	-0.96	5.34	1.091	0.918	0	3.810	-0.39	HPV
387	55283-68-6	Ethalfluraline*, §	-1.34	-2.55	5.11	0.996	0.921	0	3.002	-2.24	N
388	29091-05-2	Dinitramine*	-1.86		4.30	0.777	0.954	0	1.175		N
389	873-32-5	o-chlorobenzonitrile*	-2.85		2.18**	0.000	1.056	0	-0.153		N
390	91-15-6	Phthalonitrile*	-3.33		0.99	0.000	1.046	0	-0.691		HPV
391	140-29-4	Benzyl cyanide*	-2.94		1.56	0.311	1.010	0	0.106		HPV
392	23950-58-5	Propyzamide*, §	-2.04	-2.67	3.43	0.771	1.002	0	-0.516	-2.38	N
393	51218-45-2	Metolachlor*, §	-2.60	-0.96	3.13	0.794	0.899	0	0.706	-0.40	HPV
394	23184-66-9	Butachlor/N-(butoxymethyl)-2-chloro-2',6'-diethylacetanilide*, §	-2.31	-0.80	4.50	0.880	0.831	0	2.498	-0.21	N
395	51218-49-6	Pretilachlor/2-chloro-2',6'-diethyl-N-(2-propoxyethyl)acetanilide <Pretilachlor>*, §	-2.25	-2.22	4.08	0.821	0.889	0	1.567	-1.86	N
396	93-68-5	o-acetoacetotoluidide*	-2.50		0.99**	1.075	0.970	0	0.133		HPV
397	57837-19-1	Metalaxyl/methyl-(2-methoxyacetyl)-N-(2,6-xylyl)-DL-alaninate*, §	-3.47	-2.08	1.65	0.718	0.906	0	-2.118	-1.70	N
398	3766-81-2	Fenobucarb/2-sec-butylphenyl N-methylcarbamate*, §	-2.30	-1.47	2.78	0.929	0.924	0	1.252	-0.99	N
399	114-26-1	Propoxur*	-2.72		1.52	1.066	0.914	0	0.178		HPV
400	34123-59-6	Isoproturon/1,1-dimethyl-3-(8-isopropylphenyl)-urea*, §	-1.94	-2.27	2.87	1.053	0.971	0	0.504	-1.92	HPV
401	330-54-1	Diuron/1-(3,4 dichlorophenyl)-3,3 dimethyl urea*, §	-1.97	-1.83	2.68	0.938	1.007	0	-0.127	-1.41	HPV
402	5329 12 4	2,4,6-trichlorophenylhydrazine*	0.72		2.73**	3.141	1.050	0	-0.794		HPV
403	108-98-5	Benzenethiol*	-2.84		2.52	0.000	1.035	0	0.274		N
404	28249-77-6	Thiobencarb*, §	-2.26	-1.76	3.40	0.655	0.953	0	1.165	-1.32	N
405	88-19-7	o-toluenesulfonamide*	-3.41		0.84	0.000	1.013	0	0.531		HPV
406	63-74-1	Sulphanilamide*	-3.67		-0.62	0.123	1.033	0	-0.264		N
407	98-59-9	4-toluenesulfonyl chloride (p-Toluene sulfonyl chloride stabilised)*	-1.65		3.49**	0.810	1.051	0	-0.430		N
408	121-03-9	4-nitrotoluene-2-sulphonic acid*	-3.99		-0.80**	0.019	1.026	0	-1.011		HPV
409	15318-45-3	Thiamphenicol*	-2.97		-0.33**	1.118	0.956	0	-0.289		N
410	122-14-5	Fenitrothion*	-1.83		3.30	1.030	0.972	0	0.505		N
411	26087-47-8	Iprobenfos*	-2.68		3.34	0.619	0.904	0	0.919		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
412	110-86-1	Pyridine*	-3.59		0.65	0.000	0.999	0	0.316		HPV
413	10500-57-9	5,6,7,8-tetrahydroquinoline*	-2.82		2.77**	0.000	1.009	0	1.077		N
414	100-43-6	4-vinylpyridine*	-3.26		1.71**	0.000	0.999	0	0.528		N
415	100-69-6	2-vinylpyridine*	-3.25		1.54	0.000	1.001	0	0.844		HPV
416	462-08-8	3-aminopyridine*	-3.94		0.11	0.000	0.993	0	-0.806		N
417	504-24-5	4-aminopyridine*, §	-3.86	-1.94	0.32	0.000	0.995	0	-0.689	-1.53	N
418	504-29-0	2-aminopyridine*	-3.74		0.48	0.000	0.996	0	-0.254		N
419	1007-28-9	Atrazine-deisopropyl*	-1.92		1.15	1.297	1.012	0	0.183		N
420	122-34-9	Simazine*, §	-1.37	-2.51	2.18	1.434	1.012	0	0.915	-2.20	HPV
421	1912-24-9	Atrazine*, §	-1.30	-0.33	2.61	1.482	0.998	0	0.785	0.34	HPV
422	5915-41-3	Terbuthylazine*, §	-1.19	-2.75	3.21	1.512	0.983	0	0.878	-2.48	N
423	33693-04-8	Terbumeton*, §	-2.35	-0.28	3.10	0.697	0.951	0	0.868	0.39	N
424	1014-70-6	Simetryn*, §	-1.21	-2.42	2.80	1.570	0.993	0	0.590	-2.09	N
425	886-50-0	Terbutryn*, §	-0.71	-2.80	3.74	1.872	0.966	0	1.085	-2.53	N
426	28159-98-0	Irgarol 1051/2-methylthio-4-tert-butylamino-6-cyclopropylamino-s-triazine*, §	-0.60	-0.47	4.07**	1.869	0.975	0	0.881	0.17	N
427	51-21-8	5-fluorouracil*, §	-3.84	-2.60	-0.89	0.000	1.029	0	0.219	-2.30	N
428	21087-64-9	Metribuzin*, §	-0.96	-1.90	1.70	2.107	0.977	0	0.973	-1.49	N
429	110-02-1	Thiophene*	-2.87		1.81	0.000	1.054	0	0.546		N
430	443-48-1	Metronidazole*	-4.12		-0.02	0.012	0.956	0	-0.192		N
431	61-82-5	3-amino-1,2,4-triazole*	-4.26		-0.97	0.000	0.992	0	-0.863		HPV
432	92-52-4	Biphenyl*, §	-1.52	-1.58	4.01	0.906	1.017	0	0.203	-1.12	HPV
433	5707-44-8	4-ethyl-1,1'-biphenyl*	-1.11		4.80**	1.113	1.001	0	0.622		N
434	90-43-7	2-phenylphenol*, §	-1.98	-1.48	3.09	0.881	1.005	0	-0.418	-1.00	HPV
435	92-69-3	p-phenylphenol*, §	-1.62	-1.76	3.20	0.985	1.016	1	-0.219	-1.33	N
436	92-88-6	4,4'-dihydroxy-biphenyl*	-1.85		2.80**	1.056	1.015	0	-0.788		HPV
437	119-93-7	o-tolidine*	-1.84		2.34	1.104	1.002	2	-1.156		N
438	91-94-1	3,3'-dichlorobenzidine*, §	-1.32	-1.15	3.51	1.186	1.033	2	-1.806	-0.62	HPV
439	58-14-0	Pyrimethamine*, §	-2.42	-1.78	2.69	0.840	0.965	0	-0.531	-1.35	N
440	91-76-9	2,4-diamino-6-phenyl-s-triazine*	-2.14		1.36	1.079	1.022	0	-0.491		HPV
441	1698-60-8	Chloridazon/5-amino-4-chloro-2-pheny l-3(2H)-pyridazinone (CHD)*	-1.58		1.14	1.430	1.030	1	-0.078		HPV
442	51963-82-7	Benzenamine,2,5-diethoxy-4-(4-morpholinyl)*	-2.85		2.01**	1.054	0.903	0	-1.008		N
443	32809-16-8	Procymidone/N-(3,5-dichlorophenyl)-1,2-dimethylcyclopropane-1,2-dicarboximide*	-1.77		3.08	0.760	1.037	0	0.501		N
444	18854-01-8	Isoxathion*	-2.12		3.73	0.849	0.967	0	-0.653		N
445	19666-30-9	Oxadiazon*, §	-1.61	-1.84	4.80	1.129	0.934	0	0.134	-1.42	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
446	147-94-4	Cytarabine*	-4.08		-2.46**	0.929	0.938	0	-1.543		N
447	95058-81-4	Gemcitabine*	-3.61		-2.01**	0.904	0.961	0	-0.051		N
448	96-09-3	Styrene-7,8-oxide*	-3.21		1.61	0.000	1.016	0	0.315		N
449	901-44-0	2,2-bis[4-(2-hydroxyethoxy)phenyl]propane*	-2.19		3.22	0.775	0.946	2	-0.102		N
450	599-64-4	4-(α,α-dimethylbenzyl)phenol*	-1.61		4.12**	0.854	0.968	2	0.534		HPV
451	620-92-8	4,4'-dihydroxydiphenylmethane*	-1.53		2.91	1.000	0.982	4	0.091		N
452	79-94-7	Tetrabromobisphenol A*, §	-0.27	-1.62	7.20**	1.305	1.004	0	0.172	-1.16	HPV
453	101-77-9	4,4'-methylenedianiline*	-2.21		1.59	0.905	0.977	4	-1.050		HPV
454	101-84-8	Diphenyl ether*, §	-1.45	-1.12	4.21	1.117	0.988	0	-0.003	-0.58	HPV
455	3380-34-5	Triclosan*, §	-1.18	-0.80	4.76	1.053	1.010	1	-0.416	-0.21	N
456	101-80-4	4,4'-diaminodiphenyl ether*	-2.18		1.36	1.153	0.981	4	-2.490		HPV
457	42874-03-3	Oxyfluorfen*, §	-1.45	-0.24	4.73	0.892	0.936	0	3.084	0.44	N
458	103-50-4	Dibenzyl ether*	-2.15		3.31	0.980	0.937	0	0.307		N
459	49562-28-9	Fenofibrate*	-2.25		5.19**	0.574	0.910	0	0.383		N
460	31127-54-5	2,3,4,4'-tetrahydroxybenzophenon*	-1.49		2.42**	1.127	0.995	3	0.499		N
461	122-39-4	Diphenylamine*, §	-1.60	-1.35	3.50	1.142	0.997	0	-0.298	-0.85	HPV
462	620-93-9	Di-p-tolylamine*	-1.63		4.39**	0.900	0.991	0	-0.056		N
463	13684-63-4	Phenmedipham*, §	-1.72	-1.13	3.59	0.841	0.996	2	-0.382	-0.59	HPV
464	101-20-2	3,4,4'-trichlorodiphenylurea*	-1.01		4.90**	1.149	1.043	0	-1.009		N
465	102-06-7	1,3-diphenylguanidine*	-2.16		2.89**	0.922	0.993	0	-1.021		HPV
466	97-39-2	N,N'-bis(2-methylphenyl)guanidine*	-2.30		3.99**	0.674	0.957	0	-0.590		N
467	122-66-7	Hydrazobenzene*, §	-1.94	-1.78	2.94	1.168	0.971	0	-0.536	-1.35	N
468	60-09-3	Aniline, p-(phenylazo)-（p-Aminoazobenzene）*	-1.34		3.41	1.194	1.014	2	-0.978		N
469	80-09-1	Bis(4-hydroxyphenyl)sulfone*	-1.93		1.65**	0.778	1.011	4	0.209		HPV
470	30171-80-3	Dibromocresyl glycidyl ether*	-1.69		2.77**	1.214	0.995	0	0.023		N
471	68-35-9	Sulfadiazine*, §	-3.47	-3.09	-0.09	0.612	0.978	0	-1.217	-2.87	N
472	57-68-1	Sulfamethazine/Sulfadimidine*, §	-3.69	-2.83	0.19	0.477	0.954	0	-1.174	-2.57	N
473	122-11-2	Sulfadimethoxine*, §	-2.12	-3.11	1.63	1.451	0.946	1	-1.063	-2.90	N
474	64902-72-3	Chlorsulfuron*	-2.86		2.00	0.803	0.959	0	-1.702		N
475	77732-09-3	Oxadixyl*, §	-3.38	-1.45	0.80	0.835	0.916	0	-1.271	-0.96	N
476	90-12-0	1-methylnaphthalene*, §	-2.52	-1.53	3.87	0.000	1.020	0	0.527	-1.06	N
477	91-57-6	2-methylnaphthalene*, §	-1.70	-1.50	3.86	0.694	1.030	0	0.327	-1.02	N
478	573-98-8	1,2-dimethylnaphthalene*	-1.80		4.31	0.548	1.011	0	0.754		N
479	575-41-7	1,3-dimethylnaphthalene*	-1.74		4.42	0.548	1.019	0	0.599		N
480	582-16-1	2,7-dimethylnaphthalene*	-0.86		4.26**	1.342	1.028	0	0.427		N
481	29253-36-9	Isopropylnaphthalene*	-1.83		4.63**	0.532	0.991	0	0.969		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
482	525-66-6	Propranolol*, §	-2.03	-1.69	3.48	0.730	0.923	4	0.318	-1.24	N
483	135-19-3	B-Naphthol*, §	-2.44	-1.82	2.70	0.395	1.033	0	-0.307	-1.39	HPV
484	91-59-8	B-Naphthylamine*	-2.50		2.28	0.539	1.031	0	-0.980		N
485	479-27-6	1,8-naphthylenediamine*	-3.39		1.78	0.000	1.011	0	-1.015		N
486	2243-62-1	1,5-naphthalenediamine*	-3.68		0.89	0.000	1.016	0	-1.512		HPV
487	92-70-6	3-hydroxy-2-naphthoic acid*	-2.04		3.05	0.814	1.034	0	-1.553		HPV
488	58-27-5	2-methyl-1,4-naphthoquinone*, §	-1.88	-1.02	2.20	0.725	1.046	0	1.317	-0.46	N
489	117-80-6	2,3-dichloro-1,4-naphthoquinone*	1.62		2.65**	3.653	1.090	0	-0.263		N
90	1785-65-5	2-acetoxy-1,4-naphthoquinone*	-1.97		1.26**	1.096	1.039	0	-0.083		N
491	83-72-7	2-hydroxy-1,4-naphthoquinone*	-3.20		1.38	0.010	1.048	0	-0.711		N
492	91-22-5	Quinoline*, §	-2.98	-2.73	2.03	0.000	1.038	0	0.056	-2.45	HPV
493	91-53-2	6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline*, §	-1.73	-1.45	3.87**	1.009	0.959	0	0.884	-0.96	N
494	148-24-3	8-Hydroxyquinoline*, §	-3.06	-1.90	2.02	0.000	1.030	0	-0.079	-1.49	N
495	22720-75-8	1-benzo[b]thien-2-ylethan-1-one*	-1.76		2.67**	0.701	1.064	0	0.586		N
496	95-31-8	N-(tert-butyl)-2-benzothiazolylsulfenamide*	-1.37		4.67	0.836	1.004	0	1.094		HPV
497	95-33-0	N-cyclohexyl-2-benzothiazolylsulfenamide*	-1.32		3.47**	1.172	1.004	0	1.144		HPV
498	149-30-4	2-mercaptobenzothiazole*, §	-2.59	-1.82	2.42	0.000	1.096	0	-0.573	-1.39	HPV
499	11070-44-3	Tetrahydromethylphthalic anhydride*	-3.09		2.54**	0.000	1.003	0	-0.042		HPV
500	85-44-9	Phthalic anhydride*	-3.09		1.60	0.000	1.066	0	-1.114		HPV
501	117-08-8	Tetrachlorophthalic anhydride*	-1.85		4.65**	0.000	1.154	0	-2.421		HPV
502	83-32-9	Acenaphthene*, §	-2.29	-2.51	3.92	0.000	1.055	0	0.399	-2.20	HPV
503	85-01-8	Phenanthrene*, §	-1.23	-0.97	4.46	0.928	1.044	0	-0.068	-0.41	N
504	86-73-7	Fluorene*	-1.17		4.18	0.898	1.058	0	0.491		N
505	193-39-5	Indeno[1,2,3-cd]pyrene*	-0.07		6.70**	1.243	1.084	0	-0.593		N
506	5522-43-0	1-nitropyrene*	-0.88		5.06	0.993	1.062	0	-0.128		N
507	1484-13-5	9-vinylcarbazole*	-1.41		4.19**	0.836	1.039	0	0.207		N
508	2222-33-5	Dibenzo[b,f]cyclohepten-1-one*	-0.99		4.32**	1.172	1.041	0	0.094		N
509	132-65-0	Dibenzothiophene*, §	-1.28	-0.95	4.38	0.680	1.082	0	-0.049	-0.38	N
510	1916-55-8	2-acetamidophenoxazin-3-one*	-2.18		0.79**	1.026	1.043	1	-1.006		N
511	1916-59-2	2-aminophenoxazin-3-one*	-1.78		0.85**	1.315	1.053	1	-1.074		N
512	92-84-2	Phenothiazine*	-1.18		4.15	0.907	1.073	0	-0.220		HPV
513	14698-29-4	Oxolinic acid*, §	-2.62	-1.82	0.94	0.828	1.020	0	-0.973	-1.40	N
514	42835-25-6	Flumequine*, §	-2.31	-2.65	1.60	0.685	1.019	0	1.028	-2.36	N
515	2439 01 2	6-methyl-1,3-dithiolo[4,5-b]quinoxalin-2-one*	0.15		3.78	2.095	1.098	0	-0.741		N
516	90-30-2	1-(N-phenylamino)-naphthalene*	-1.35		4.20	1.139	1.009	0	-0.379		HPV
517	135-88-6	N-phenyl-2-naphthylamine*	-1.56		4.38	0.965	1.005	0	-0.690		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
518	88426-33-9	Buparvaquone*	-1.00		5.86**	1.012	0.958	1	1.347		N
519	79617-96-2	Sertraline*, §	-1.17	-1.16	5.29**	0.939	0.978	1	0.937	-0.63	N
520	70458-96-7	Norfloxacin*, §	-3.14	-3.00	-1.03	0.957	0.975	0	0.241	-2.77	N
521	85721-33-1	Ciprofloxacin*, §	-2.99	-1.09	0.28	0.918	0.983	0	-1.245	-0.55	N
522	93106-60-6	Enrofloxacin*, §	-2.84	-2.15	0.70	0.992	0.954	0	-0.289	-1.78	N
523	98079-51-7	Lomefloxacin*, §	-2.72	-2.83	-0.30	0.973	0.982	0	1.325	-2.57	N
524	948-65-2	2-Phenylindole*	-1.09		3.82**	1.212	1.048	0	-0.264		N
525	115-86-6	Triphenyl phosphate*, §	-2.33	-1.04	4.59	0.655	0.939	0	-0.980	-0.49	HPV
526	27955-94-8	Phenol,4,4',4''-ethylidynetris-*	-0.35		4.38**	1.074	0.963	12	-0.378		N
527	79-57-2	Oxytetracycline*	-2.75		-0.90	0.828	0.953	5	0.909		N
528	95233-18-4	Atovaquone*	-0.76		5.87**	1.002	1.006	2	0.098		N
529	23696-28-8	Olaquindox*	-3.61		-2.13**	1.028	0.946	1	-0.822		N
530	58-89-9	Lindane (γ-HCH)*, §	-2.37	-1.24	3.72	0.000	1.085	0	-1.096	-0.72	HPV
531	75-35-4	1,1-dichloroethylene*	-2.97		2.13	0.000	1.036	0	0.057		HPV
532	67-56-1	Methanol*	-5.14		-0.77	0.000	0.839	0	-0.318		HPV
533	67-63-0	2-propanol*	-4.24		0.05	0.000	0.922	0	0.464		HPV
534	20679-58-7	Acetic acid, bromo-, 2-butene-1,4-diyl ester (Fennosan F50)*	-2.51		1.86**	0.942	0.981	0	-1.094		N
535	80-62-6	Methyl methacrylate*	-3.85		1.38	0.000	0.953	0	-0.751		HPV
536	75-64-9	t-butylamine*	-4.06		0.40	0.000	0.927	0	0.844		HPV
537	124-40-3	Dimethylamine*	-4.49		-0.38	0.000	0.899	0	0.611		HPV
538	108-18-9	Diisopropylamine*	-3.71		1.40	0.000	0.926	0	1.362		HPV
539	111-92-2	Dibutylamine*	-1.76		2.83	0.931	0.989	0	1.875		HPV
540	68-12-2	Dimethylformamide*	-4.45		-1.01	0.000	0.935	0	0.513		HPV
541	62-75-9	Dimethylnitrosamine*	-4.39		-0.57	0.000	0.932	0	0.203		N
542	55-18-5	Diethylnitrosamine*	-4.07		0.48	0.000	0.920	0	0.909		N
543	99129-21-2	Clethodim*	-1.64		4.21**	0.767	0.933	2	2.370		N
544	62-56-6	Thiourea*	-4.59		-1.08	0.000	0.944	0	-0.706		HPV
545	2212-67-1	Molinate*, §	-1.99	-1.88	3.21	0.720	0.998	0	0.849	-1.46	HPV
546	77182-82-2	Glufosinate*	-5.41		-4.49**	0.070	0.973	0	-1.710		N
547	1071-83-6	Glyphosate*	-5.08		-3.40	0.015	0.986	0	-1.664		HPV
548	126-72-7	Tris-(2,3-dibromopropyl)-phoshate*	-2.04		4.29	0.918	0.908	0	0.963		N
549	115-29-7	Endosulfan*, §	0.13	-0.59	3.83	2.092	1.074	0	0.069	0.04	HPV
550	131-11-3	Dimethyl phthalate*, §	-3.28	-2.54	1.60	0.519	0.960	0	-1.613	-2.23	HPV
551	644-35-9	2-n-propylphenol*	-2.45		2.93	0.535	0.980	0	0.497		N
552	98-54-4	p-tert-butylphenol*	-2.44		3.31	0.528	0.966	0	0.514		HPV
553	609-19-8	3,4,5-trichlorophenol*, §	-2.22	-1.68	4.01	0.000	1.089	0	-0.780	-1.23	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
554	935-95-5	2,3,5,6-tetrachlorophenol*	-2.11		3.88	0.000	1.116	0	-0.978		N
555	4901-51-3	2,3,4,5-tetrachlorophenol*	-2.03		4.21	0.000	1.116	0	-1.091		N
556	2460-49-3	4,5-dichloroguaiacol*	0.59		3.26	3.087	1.018	0	-0.836		N
557	2668-24-8	4,5,6-trichloroguaiacol*	0.64		3.72	2.982	1.038	0	-1.485		N
558	57057-83-7	3,4,5-trichloroguaiacol*	0.76		3.77	2.982	1.030	0	-0.369		N
559	2539-17-5	Tetrachloroguaiacol*, §	1.10	-1.71	4.59	3.012	1.050	0	-0.753	-1.27	N
560	2539-26-6	3,4,5-trichloro-2,6-dimethoxyphenol*	-0.24		3.74	2.301	1.001	0	-0.585		N
561	100-02-7	4-nitrophenol*, §	-3.28	-2.32	1.91	0.031	1.014	0	-0.873	-1.98	HPV
562	1689-84-5	Bromoxynil*, §	-2.15	-1.50	3.39**	0.153	1.095	0	-0.586	-1.02	N
563	108-46-3	Resorcinol*	-3.67		0.80	0.000	1.000	0	-0.598		HPV
564	120-80-9	Catechol*	-3.69		0.88	0.000	0.998	0	-0.743		HPV
565	615-67-8	Chlorohydroquinone*	-3.30		1.40	0.000	1.028	0	-0.471		N
566	95-88-5	4-chlororesorcinol*	-3.20		1.80	0.000	1.027	0	-0.495		N
567	2138-22-9	4-chlorocatechol*	-3.29		1.68**	0.000	1.026	0	-0.800		N
568	3428-24-8	4,5-dichlorocatechol*	-2.95		2.32**	0.000	1.053	0	-0.946		N
569	3978-67-4	3,4-dichlorocatechol*	-2.97		2.32**	0.000	1.051	0	-0.978		N
570	13673-92-2	3,5-dichlorocatechol*	-2.97		2.32**	0.000	1.052	0	-1.040		N
571	137-19-9	4,6-dichlororesorcinol*	-2.82		2.32**	0.000	1.054	0	-0.119		N
572	56961-20-7	3,4,5-trichlorocatechol*	-2.47		3.71	0.000	1.077	0	-1.419		N
573	32139-72-3	3,4,6-trichlorocatechol*	-2.52		3.60	0.000	1.078	0	-1.547		N
574	1198-55-6	Tetrachlorocatechol*	-2.23		4.29	0.000	1.102	0	-1.968		N
575	87-87-6	Tetrachlorohydroquinone*	-2.33		3.61**	0.000	1.104	0	-1.431		N
576	87-66-1	1,2,3-trihydroxybenzene*	-3.65		0.97**	0.000	0.997	0	-0.632		N
577	99-55-8	2-amino-4-nitrotoluene *	-3.39		1.87	0.032	1.000	0	-0.940		N
578	119-32-4	4-amino-2-nitrotoluene *	-3.34		2.02**	0.000	1.001	0	-0.706		N
579	603-83-8	2-amino-6-nitrotoluene*, §	-3.34	-1.10	2.02**	0.000	0.994	0	-0.372	-0.56	N
580	19406-51-0	4-amino-2,6-dinitrotoluene*, §	-3.38	-1.08	1.84**	0.000	1.006	0	-0.840	-0.54	N
581	35572-78-2	2-amino-4,6-dinitrotoluene*, §	-3.37	-1.54	1.84**	0.036	1.005	0	-1.023	-1.07	N
582	823-40-5	2,6-diaminotoluene *	-4.07		0.16**	0.000	0.981	0	-1.213		HPV
583	6629-29-4	2,4-diamino-6-nitrotoluene*	-3.76		0.55**	0.000	0.994	0	-0.416		N
584	59229-75-3	2,6-diamino-4-nitrotoluene*	-3.90		0.55**	0.034	0.994	0	-1.681		N
585	56-75-7	Chloramphenicol*	-2.83		1.14	0.941	0.954	0	-0.649		N
586	121-14-2	2,4-dinitrotoluene*, §	-3.28	-2.58	1.98	0.034	1.013	0	-1.013	-2.28	HPV
587	118-96-7	2,4,6-trinitrotoluene*, §	-3.38	-1.94	1.60	0.038	1.016	0	-1.135	-1.54	HPV
588	1194-65-6	2,6-dichlorobenzonitrile*, §	-2.48	-1.53	2.74	0.000	1.084	0	0.057	-1.06	N
589	1897-45-6	Tetrachloroisophthalonitrile*	-2.07		3.05	0.000	1.147	0	-0.531		HPV
590	34256-82-1	Acetochlor*, §	-2.25	-1.01	3.03	1.016	0.905	0	1.316	-0.45	HPV

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
591	15545-48-9	Chlorotoluron*, §	-2.07	-2.39	2.41	0.949	0.996	0	0.138	-2.06	HPV
592	23564-05-8	Dimethyl 4,4'-(o-phenylente) bis(3-thioa(lophanate)*	-3.02		1.40	0.899	0.922	0	-0.691		N
593	57-67-0	Sulfaguanidine*	-3.63		-1.22	0.578	1.010	0	-1.352		N
594	73231-34-2	Florfenicol*, §	-2.87	-3.17	-0.04**	1.080	0.950	0	0.388	-2.97	N
595	64249-01-0	Anilofos*	-1.55		3.81	1.449	0.920	0	0.602		N
596	22224-92-6	Fenamiphos*, §	-2.26	-3.10	3.23	0.774	0.941	0	1.102	-2.88	HPV
597	69377-81-7	Fluroxypyr*	-2.03		2.20	1.154	1.012	1	-2.276		N
598	21725-46-2	Cyanazine*, §	-1.78	-2.37	2.22	1.444	0.948	0	0.850	-2.04	N
599	834-12-8	Ametryn*, §	-1.11	-2.28	2.98	1.740	0.968	0	0.773	-1.93	HPV
600	7287-19-6	Prometryn*, §	-0.82	-1.60	3.51	1.872	0.961	0	0.998	-1.14	N
601	59-87-0	Nitrofurazone*, §	-2.74	-2.47	0.23	1.064	1.008	0	-1.696	-2.15	N
602	34014-18-1	Tebuthiuron*, §	-1.58	-1.64	1.79	1.487	0.985	0	0.989	-1.19	N
603	119-12-0	Pyridaphenthion*	-2.50		3.20	0.874	0.915	0	-0.057		N
604	24096-53-5	N-(3,5-dichlorophenyl)succinidide*	-0.36		1.40	2.396	1.055	0	-0.088		N
605	36734-19-7	3-(3,5-dichlorophenyl)-N-isopropyl-2,4-dioxoimidazolidine-l-carboxamide*	-1.51		3.00	0.804	1.010	3	1.094		HPV
606	39807-15-3	Oxadiargyl*	-1.21		3.95	1.378	1.000	0	-0.446		N
607	51338-27-3	Diclofop-methyl/2-[4-(2,4-dichlorophenoxy)]-phenoxy propionate methyl ester*, §	-1.95	-2.06	4.62	0.991	0.947	0	-1.361	-1.67	HPV
608	40843-25-2	Diclofop-P*	-2.05		4.58	0.891	0.967	0	-2.117		N
609	40843-73-0	4-(2,4-dichlorophenoxy)-phenol*, §	-1.18	-2.23	4.02**	1.091	1.008	3	-0.713	-1.87	HPV
610	68359-37-5	Beta-cyfluthrin*, §	-1.93	-1.92	5.95	0.970	0.878	0	-0.504	-1.51	N
611	54910-89-3	Fluoxetine*	-1.19		4.05	1.104	0.957	1	2.792		N
612	22071-15-4	Ketoprofen*, §	-2.02	-2.61	3.12	0.955	0.956	2	-0.596	-2.31	N
613	85-68-7	Butylbenzyl phthalate*, §	-2.43	-1.30	4.73	0.951	0.855	0	-0.364	-0.79	HPV
614	71626-11-4	R-(－)-benalaxyl/Rac-benalaxyl/S-(+)-benalaxyl*	-2.75		3.40	0.765	0.895	0	-0.390		N
615	126833-17-8	Fenhexamid*, §	-1.13	-3.11	3.51	1.299	0.969	3	0.737	-2.89	N
616	72619-32-0	Haloxyfop-R*, §	-1.71	-0.72	4.05	1.095	0.944	0	0.717	-0.12	N
617	83066-88-0	Fluazifop-p*	-1.83		3.58**	0.948	0.960	0	1.202		N
618	83055-99-6	Bensulfuron-methyl*, §	-3.23	-0.58	2.18	1.227	0.814	1	-2.021	0.05	N
619	90982-32-4	Chlorimuron-ethyl*, §	-2.87	-3.05	2.50	1.132	0.900	0	-2.578	-2.83	N
620	111991-09-4	Nicosulfuron*, §	-3.47	-2.80	0.01	0.936	0.908	0	-0.774	-2.53	N
621	136849-15-5	Cyclosulfamuron*, §	-2.58	-1.58	2.05	1.314	0.820	3	0.194	-1.12	N
622	74223-64-6	Metsulfuron-methyl*, §	-3.33	-1.94	2.20	0.718	0.912	0	-2.514	-1.54	N
623	106040-48-6	Tribenuron*	-2.99		1.70**	0.847	0.952	0	-1.985		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
624	111353-84-5	Ethametsulfuron*	-3.19		1.59**	0.983	0.929	1	-3.951		N
625	79319-85-0	N,N'-methylene-di(2-amino-5-mercapto-1,3,4-thiodiazole)*	-2.81		-2.12**	1.172	1.059	0	-0.853		N
626	93697-74-6	Pyrazosulfuron-ethyl*	-2.94		1.30	1.222	0.900	0	-1.432		N
627	84087-01-4	Quinclorac*, §	-1.65	-3.14	2.97**	0.991	1.064	0	-1.470	-2.93	N
628	52316-55-9	Carbendazim*	-2.07		1.55**	1.118	1.029	0	-1.025		N
629	17804-35-2	Methyl-l-(butylcarbamoyl)-2-benzimidazole carbamate*, §	-2.04	-2.06	2.12	1.163	0.962	1	0.021	-1.67	N
630	18691-97-9	Methabenzthiazuron*	-1.83		2.64	0.970	1.018	0	0.210		HPV
631	25059-80-7	Benazolin-ethyl*, §	-2.12	-2.96	2.50	1.138	0.959	0	-0.176	-2.72	N
632	260-94-6	Acridine*, §	-1.13	-1.57	3.40	1.232	1.054	0	-0.157	-1.11	N
633	59-40-5	Sulfaquinoxaline*, §	-2.70	-2.89	1.68	0.868	0.972	0	-1.031	-2.64	N
634	94051-08-8	Quizalofop-p*	-2.33		3.57**	0.856	0.955	0	-1.272		N
635	73250-68-7	Mefenacet*, §	-1.96	-1.29	3.23	1.076	0.955	0	0.167	-0.78	HPV
636	95617-09-7	Fenoxaprop*	-2.00		4.17**	0.874	0.996	0	-2.161		N
637	98967-40-9	Flumetsulam*, §	-1.60	-0.76	1.50**	1.170	1.050	0	0.910	-0.16	N
638	139-91-3	Furaltadone*	-2.85		0.25**	0.902	0.990	0	-0.429		N
639	87818-31-3	Cinmethylin*, §	-1.56	-2.13	4.62	1.006	0.922	0	2.285	-1.76	N
640	125401-75-4	Bispyribac*	-2.61		1.25**	1.391	0.873	2	-0.730		N
641	564-25-0	Deoxytetracycline*	-2.74		-0.02	0.837	0.962	6	-1.883		N
642	100986-85-4	Levofloxacin*	-2.91		-0.20**	0.983	0.975	0	0.065		N
643	41083-11-8	Azocy-clotin/1-(tricyclohexylstannyl)-1H-1,2,4-triazole*	-1.42		5.30	0.257	1.039	0	2.503		N
644	76-87-9	Fentin hydroxide*	-1.92		3.53	0.225	1.098	0	0.057		N
645	13121-70-5	Cyhexatin*	-1.18		6.63**	0.237	1.018	0	2.663		N
646	4474-24-2	Acid blue 80*	-0.64		6.77**	0.919	0.936	10	-3.012		N
647	41859-67-0	Bezafibrate*, §	-2.76	-2.25	4.25**	0.772	0.842	1	-0.519	-1.90	N
648	13171-00-1	Celestolide*	-0.21		5.93**	1.560	0.967	0	2.681		N
649	15307-86-5	Diclofenac*, §	-0.82	-2.10	4.51	1.438	0.978	4	-1.294	-1.72	N
650	54739-18-3	Fluvoxamine*	-1.69		3.09**	1.159	0.878	3	2.906		N
651	1222-05-5	Galaxolide*	-0.50		5.90	1.328	0.977	0	2.117		HPV
652	81-14-1	Musk ketone*	-2.05		4.30	0.349	0.958	0	2.998		N
653	81-15-2	Musk xylene*	-2.23		4.45**	0.187	0.967	0	2.364		N
654	61869-08-7	Paroxetine*	-1.56		4.74**	0.965	0.920	0	2.464		N
655	1506-02-1	Tonalide*	-0.98		6.35**	0.889	0.960	0	2.125		HPV
656	55268-75-2	Cefuroxime*	-3.48		-0.16	1.072	0.944	0	-3.192		N
657	26787-78-0	Amoxicillin*	-2.53		0.87	0.981	0.943	2	0.564		N
658	15686-71-2	Cephalexin*	-2.53		0.65	0.913	1.013	1	-0.885		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
659	2022-85-7	5-fluorocytosine*	-4.19		-1.41**	0.000	1.016	0	-0.630		N
660	16110-51-3	Cromolyn*	-3.68		1.55**	0.991	0.819	1	-2.289		N
661	73-22-3	L-tryptophan*	-3.25		-1.06	0.950	0.981	1	-1.371		N
662	59-05-2	Methotrexate*	-3.84		-1.85	1.155	0.864	4	-2.488		N
663	51-52-5	Propylthiouracil*, §	-1.63	-2.15	0.98**	1.726	1.001	0	-0.368	-1.78	N
664	60-80-0	Antipyrine*	-2.60		0.38	1.031	0.994	0	-0.183		HPV
665	87-08-1	Phenoxymethylpenicillinic Acid*	-2.48		2.09	0.851	0.980	0	-0.552		N
666	64544-07-6	Cefuroxime Axetil*	-2.81		0.89	1.210	0.944	1	-2.465		N
667	33419-42-0	Etoposide*	-2.87		0.60	1.012	0.916	4	-1.736		N
668	51481-61-9	Cimetidine*	-3.08		0.40	1.092	0.894	0	0.599		N
669	94-20-2	Chlorpropamide*	-2.27		2.27	1.159	0.926	1	-0.197		N
670	3930-20-9	Sotalol*, §	-2.68	-3.23	0.24	1.098	0.956	0	0.733	-3.04	N
671	58-93-5	Hydrochlorothiazide*	-2.34		-0.07	0.810	1.068	0	0.841		HPV
672	1156-19-0	Tolazamide*	-2.31		2.69	0.943	0.947	1	-0.563		HPV
673	50-23-7	Hydrocortisone*	-1.92		1.61	0.901	0.957	4	1.827		N
674	50-24-8	Prednisolone*	-2.00		1.62	0.896	0.967	4	0.835		N
675	51-34-3	Scopolamine*	-2.75		0.98	0.998	0.954	0	-0.292		N
676	26839-75-8	Timolol*	-3.28		1.83	0.783	0.815	0	2.436		N
677	37350-58-6	Metoprolol*, §	-2.31	-0.71	1.88	0.730	0.964	2	1.134	-0.11	N
678	1091-85-6	Dansylglycine*	-1.55		1.44**	1.361	0.975	2	1.771		N
679	137-58-6	Lidocaine*	-2.89		2.44	0.673	0.900	0	0.912		N
680	83-43-2	Methylprednisolone*	-1.81		1.82**	1.004	0.968	4	0.834		N
681	64-77-7	Tolbutamide*	-2.71		2.34	0.851	0.892	1	0.596		N
682	526-08-9	Sulfaphenazole*	-2.85		1.52	0.904	0.940	1	-1.357		N
683	37517-30-9	Acebutolol*, §	-2.40	-2.50	1.71	0.899	0.952	1	0.842	-2.18	N
684	59-46-1	Procaine*	-2.59		2.14	0.985	0.952	0	-1.165		N
685	63590-64-7	Terazosin*	-2.54		1.47**	0.973	0.966	0	-0.079		N
686	6452-71-7	Oxprenolol*	-2.71		2.10	0.728	0.885	2	1.542		N
687	84057-84-1	Lamotrigine*	-1.18		2.57	1.356	1.021	2	0.108		N
688	4205-90-7	Clonidine*	-1.02		1.59	1.943	1.023	0	0.004		N
689	54-31-9	Furosemide*	-2.22		2.03	0.975	0.992	1	-0.948		HPV
690	66357-35-5	Ranitidine*, §	-3.47	-3.47	0.27	1.430	0.794	1	-0.720	-3.31	N
691	7689-03-4	Camptothecin*	-1.87		1.74	0.938	1.028	2	-0.048		N
692	34841-39-9	Bupropion*	-1.24		3.85**	1.525	0.946	0	0.852		N
693	103628-46-2	Sumatriptan*	-2.60		0.93	1.151	0.929	0	0.801		N
694	81-81-2	Warfarin*, §	-2.16	-2.93	2.70	0.906	0.979	0	0.078	-2.69	N
695	28395-03-1	Bumetanide*	-1.87		2.57**	1.087	0.924	5	-0.364		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
696	129-20-4	Oxyphenbutazone*	-2.03		2.72	1.064	0.941	3	-0.873		N
697	87848-99-5	Acrivastine*	-1.77		2.83**	1.171	0.944	2	0.518		N
698	57-41-0	Phenytoin*	-2.17		2.47	0.972	0.995	0	-0.766		HPV
699	13655-52-2	Alprenolol*	-2.03		3.10	0.773	0.927	3	1.278		N
700	19216-56-9	Prazosin*	-2.64		1.28**	1.005	0.978	0	-1.249		N
701	36894-69-6	Labetalol*	-2.05		3.09	0.940	0.911	2	1.390		N
702	50-33-9	Phenylbutazone*	-2.12		3.16	1.061	0.945	0	-0.187		N
703	637-07-0	Clofibrate*, §	-2.64	-2.21	3.62**	0.560	0.932	0	-0.135	-1.85	N
704	94-24-6	Tetracaine*	-1.82		3.51	1.092	0.966	0	0.006		N
705	6990 06 3	Fusidic acid*	-0.21		6.75	1.057	0.917	7	2.035		N
706	303-81-1	Novobiocin*	-1.97		2.45**	0.915	0.936	5	-0.051		N
707	99614-02-5	Ondansetron*	-1.56		3.95**	0.904	1.002	0	0.729		N
708	548-73-2	Droperidol*	-2.07		3.50	0.924	0.906	1	1.542		N
709	56-54-2	Quinidine*	-1.92		3.44	1.053	0.935	0	1.162		N
710	53-86-1	Indomethacin*	-1.66		4.27	0.963	0.962	2	-0.676		N
711	130-95-0	Quinine*	-1.91		3.44	1.053	0.930	0	1.407		N
712	599-79-1	Sulfasalazine*	-1.69		3.81**	0.883	0.971	4	-1.338		N
713	57-83-0	Progesterone*	-1.44		3.87	0.977	0.961	0	3.036		N
714	50-47-5	Desipramine*	-1.34		4.90	1.087	0.935	2	0.588		N
715	10238-21-8	Glibenclamide*	-2.60		4.79**	1.206	0.764	1	-0.313		HPV
716	50-49-7	Imipramine*	-1.39		4.80	1.066	0.948	0	1.505		N
717	65277-42-1	Ketoconazole*	-2.70		4.35	1.068	0.824	0	-1.000		N
718	58-40-2	Promazine*	-1.46		4.55	1.080	0.962	0	0.761		N
719	84625-61-6	Itraconazole*	-2.10		5.66	0.982	0.848	0	0.174		N
720	146-54-3	Triflupromazine*	-0.80		5.54	1.087	0.958	0	3.457		N
721	50-53-3	Chlorpromazine*	-1.18		5.41	1.096	0.968	0	0.634		N
722	91161-71-6	Terbinafine*	-1.12		6.00	1.125	0.928	0	1.505		N
723	23593-75-1	Clotrimazole*	-0.92		6.26**	1.180	0.970	0	0.046		N
724	**3332-27-2*** **	**N,N-Dimethyltetradecylamine N-oxide***	**-1.05**		**2.69****	**0.971**	**1.025**	**0**	**5.087**		**HPV**
725	77-58-7	Dibutyltin dilaurate*	-1.70		3.12	1.013	0.869	0	6.522		HPV
726	116-37-0	1,1'-Isopropylidenebis(p-phenyleneoxy)dipropan-2-ol*	-1.80		4.06**	0.885	0.935	2	0.586		N
727	10222-01-2	2,2-Dibromo-2-cyanoacetamide*	-3.17		0.82	0.000	1.074	0	-0.571		N
728	6021-61-0	2-[4-[(2-chloro-4-nitrophenyl)azo]-n-(2-cyanoethyl)anilino]ethyl acetate*	-1.87		4.87**	1.071	0.937	0	-1.440		N
729	50-29-3	Dichlorodiphenyltrichloroethane (DDT)*	-0.98		6.91	0.629	1.020	0	0.387		HPV
730	309-00-2	Aldrin*	2.30		6.50	3.137	1.077	0	1.600		N
731	36355-01-8	Hexabromobiphenyl*	1.08		6.39	2.218	1.110	0	-0.904		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
732	101-14-4	2,2'-dichloro-4,4'-methylendianiline*	-1.55		3.91	0.989	0.995	2	-0.971		HPV
733	31508-00-6	1,2,4-trichloro-5-(3,4-dichlorophenyl)benzene*	0.27		7.12	1.452	1.081	0	-0.584		N
734	208-96-8	Acenaphthylene*	-2.21		3.94	0.000	1.066	0	0.418		N
735	56-55-3	Benzo[a]anthracene*, §	-0.62	-0.24	5.76	1.136	1.052	0	-0.304	0.44	N
736	53-70-3	Dibenzo[a,h]anthracene*	-0.45		6.75	1.058	1.051	0	-0.402		N
737	101-55-3	1-bromo-4-phenoxybenzene*, §	-1.48	-0.62	4.94**	0.837	1.005	0	-0.174	0.00	N
738	24017-47-8	Triazophos*, §	-1.97	-1.77	3.34	0.824	0.971	0	1.131	-1.34	N
739	1461-25-2	Tetra-n-butyltin*	-0.14		9.37**	0.210	1.033	0	4.073		HPV
740	4640 01 1	Methyl triclosan*	-1.58		5.27	0.943	0.969	0	-0.705		N
741	112-18-5	N,N-dimethyldodecan-1-amine*	-0.65		5.44**	0.956	1.000	0	3.829		HPV
742	124-19-6	Nonanal*	-1.73		3.27**	0.867	0.996	0	1.443		HPV
743	526-73-8	1,2,3-trimethylbenzene*	-2.76		3.66	0.000	0.979	0	1.149		N
744	629-59-4	Tetradecane*	0.02		7.20	0.945	1.029	0	3.852		HPV
745	1120-21-4	Undecane *	-0.67		5.74**	0.918	1.009	0	3.021		HPV
746	84-65-1	9,10-Anthracenedione*	-0.94		3.39	1.236	1.065	0	0.669		HPV
747	86-74-8	Carbazole*, §	-1.36	-1.74	3.72	0.937	1.061	0	-0.336	-1.30	HPV
748	92-06-8	1,3-diphenylbenzene*	-1.00		5.52**	1.023	1.016	0	0.060		N
749	580-51-8	3-phenylphenol*	-1.83		3.23	0.912	1.014	0	-0.344		N
750	110-54-3	N-hexane*, §	-2.72	-2.05	3.90	0.000	0.962	0	1.715	-1.66	HPV
751	122-88-3	4-chlorophenoxyacetic acid*	-1.95		2.25	1.059	1.014	1	-1.142		N
752	80060-09-9	Diafenthiuron*	-0.74		6.00	1.419	0.885	2	2.205		N
753	59-30-3	Folic acid*	-4.30		-2.81**	1.052	0.860	5	-3.544		N
754	142469-14-5	Tritosulfuron*	-1.96		3.00	0.739	0.946	0	3.641		N
755	84030-86-4	Esbiothrin*	-1.97		5.52**	0.584	0.908	0	1.646		N
756	5836 10 2	Chlorpropylate*	-1.52		4.41**	1.210	0.947	0	0.231		N
757	78-34-2	Dioxathion*	-2.38		3.45**	0.676	0.926	0	1.311		N
758	957-51-7	Diphenamid*	-1.96		2.86**	1.066	0.958	0	0.808		N
759	2540-82-1	Formothion*	-3.44		1.48	0.590	0.906	0	-0.612		N
760	961-22-8	Azinphosmethyl oxon*	-2.15		0.78	1.199	1.010	0	0.108		N
761	16655-82-6	3-hydroxycarbofuran*	-2.83		0.76**	0.870	0.968	0	-0.024		N
762	3739-38-6	3-Phenoxybenzoic acid*, §	-1.45	-0.66	3.91	1.196	0.980	2	-1.178	-0.05	N
763	107-49-3	Tetraethyl pyrophosphate*	-4.21		0.45**	0.308	0.891	0	-1.044		N
764	7421-93-4	Endrin aldehyde*	1.77		4.80**	2.837	1.105	0	2.305		N
765	31972-43-7	Fenamiphos sulfoxide*, §	-2.94	-1.51	0.73**	0.868	0.933	0	0.857	-1.04	N
766	2581-34-2	3-methyl-4-nitrophenol*	-3.01		2.48	0.031	1.003	0	0.388		N
767	3761-41-9	Fenthion sulfoxide*	-2.39		1.92**	0.877	0.979	0	0.250		N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
768	3761-42-0	Fenthion sulfone*	-2.17		2.05**	1.012	0.988	0	-0.008		N
769	87237-48-7	Haloxyfop-2-ethoxyethyl*	-2.34		4.33	0.999	0.843	0	1.100		N
770	2635 10 1	Methiocarb sulfoxide*	-2.77		0.70**	0.758	0.991	0	0.338		N
771	2179-25-1	Methiocarb sulfone*	-2.71		0.84**	0.788	0.994	0	0.113		N
772	2588 03 6	Phorate sulfoxide*	-2.91		1.78	0.687	0.913	0	1.365		N
773	2588 04 7	Phorate sulfone*	-2.46		1.99	0.662	0.984	0	1.011		N
774	1942-71-8	2-(4-tert-butylphenoxy)cyclohexanol*	-1.64		4.71**	0.859	0.950	1	0.712		N
775	27304-13-8	Oxychlordane*, §	-1.20	-2.72	5.48**	0.000	1.125	0	1.753	-2.44	N
776	53380-22-6	Ethiofencarb sulfoxide*	-3.10		-0.10**	0.977	0.929	1	-0.055		N
777	53380-23-7	Ethiofencarb sulfone*	-3.01		0.01**	0.951	0.936	1	0.219		N
778	311-45-5	Ethyl paraoxon*	-2.60		1.98	0.977	0.951	0	-0.794		N
779	40020-01-7	Pyridafol*, §	-1.64	-1.60	1.87**	1.152	1.041	0	0.530	-1.14	N
780	2703-37-9	Thiometon sulfoxide*, §	-1.99	-1.29	0.73**	1.342	1.002	0	0.562	-0.78	N
781	20301-63-7	Thioometon sulfone*	-1.75		0.85**	1.335	1.042	0	0.240		N
782	95-69-2	4-chloro-2-methylaniline*	-3.19		2.27**	0.000	1.008	0	-0.460		N
783	140-38-5	(4-chlorophenyl)urea*	-2.08		1.80	1.096	1.018	0	-0.886		N
784	61898-95-1	Methyl-3-(2,2-dichlorovinyl)-2,2-dimethyl-(1-cyclopropane) carboxylate*	-2.66		3.66**	0.214	0.995	0	-0.548		N
785	1713-15-1	2,4-D-1-isobutyl ester*	-2.47		4.30**	0.522	0.954	0	-1.013		N
786	62610-77-9	Methacrifos*	-1.53		2.53**	1.641	0.968	0	-0.437		N
787	2227-13-6	Tetrasul*	-0.44		6.87**	0.969	1.058	0	-0.207		N
788	950-10-7	Mephospholan*	-2.20		1.04	1.109	0.998	0	0.540		N
789	1214-39-7	6-Benzyladenine*	-2.40		1.57	1.070	0.951	1	-0.215		N
790	120923-37-7	Amidosulfuron*, §	-2.70	-3.38	1.63	1.103	0.931	0	-0.971	-3.21	N
791	120162-55-2	Azimsulfuron*, §	-3.21	-0.42	0.65	1.233	0.895	0	-1.982	0.23	N
792	120-23-0	Bnoa; 2-naphthyloxyacetic acid*	-1.75		2.53	1.023	1.025	2	-1.302		N
793	41483-43-6	Bupirimate*, §	-1.96	-1.87	2.70	1.198	0.926	1	0.829	-1.45	N
794	55285-14-8	Carbosulfan*, §	-2.09	-0.65	5.57**	0.773	0.831	0	2.795	-0.04	N
795	1134-23-2	Cycloate*	-1.79		3.88	0.839	0.978	0	0.941		HPV
796	13684-56-5	Desmedipham*, §	-1.48	-1.13	3.39	1.100	0.999	2	-0.505	-0.59	N
797	3347-22-6	Dithianon*, §	-0.63	-2.03	2.84	1.229	1.121	0	1.335	-1.64	N
798	126801-58-9	Ethoxysulfuron*, §	-2.25	-3.33	2.89	1.282	0.866	1	0.521	-3.15	N
799	61213-25-0	Fluorochloridone*	-1.11		3.36	0.953	1.036	0	3.038		N
800	77-06-5	Gibberellic acid*, §	-3.03	-2.18	0.24	0.693	1.004	1	-1.452	-1.81	N
801	10004-44-1	Hymexazol*, §	-3.77	-1.86	0.46	0.000	1.014	0	-1.218	-1.44	N
802	81405-85-8	Imazamethabenz-methyl*	-2.40		1.68	1.131	0.950	0	-0.062		N
803	140923-17-7	Iprovalicarb*, §	-1.97	-0.82	3.33**	1.265	0.885	1	0.865	-0.23	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
804	123-33-1	Maleic hydrazide*, §	-4.21	-3.18	-0.84	0.000	1.018	0	-1.929	-2.98	N
805	133408-50-1	Metominostrobin*	-2.16		2.32	1.145	0.919	2	0.180		N
806	2310-17-0	Phosalone*	-2.01		4.38	0.695	0.988	0	-0.906		N
807	90717-03-6	Quinmerac*, §	-2.37	-2.14	0.78	0.982	1.039	0	-1.018	-1.77	N
808	111872-58-3	Halfenprox*, §	-1.43	-2.37	8.35**	0.903	0.850	0	0.099	-2.04	N
809	90035-08-8	Flocoumafen*	0.28		8.61**	0.913	0.988	2	3.441		N
810	65731-84-2	Beta cypermethrin*	-2.17		6.05	0.968	0.857	0	-1.358		N
811	56073-10-0	Brodifacoum*	-0.17		8.50	0.957	0.990	2	0.190		N
812	1469-48-3	Cis-1,2,3,6-tetrahydrophthalimide*	-3.61		0.30**	0.000	1.025	0	-0.318		N
813	6515-38-4	3,5,6-trichloro-2-pyridinol*	-2.31		3.21	0.000	1.107	0	-0.707		N
814	1031-07-8	Endosulfan sulfate*, §	0.19	-1.14	3.66	2.082	1.076	0	0.844	-0.60	N
815	120068-36-2	Fipronil sulfone*	-0.43		4.42**	1.292	1.051	0	2.349		N
816	120067-83-6	Fipronil sulfide*	-0.39		4.82**	1.163	1.046	0	3.091		N
817	1689-83-4	Ioxynil*	-2.25		3.43	0.040	1.113	0	-1.295		N
818	1646-87-3	Aldicarb-sulfoxide*	-3.31		-0.78**	0.846	0.962	0	0.041		N
819	1646-88-4	Aldicarb-sulfone*	-3.27		-0.57	0.822	0.965	0	-0.017		N
820	3032-40-4	Fluometuron desmethyl*	-1.87		2.14	0.807	1.008	0	2.571		N
821	1570-64-5	2-methyl-4-chlorophenol*, §	-2.92	-1.02	2.78	0.000	1.018	0	-0.081	-0.46	HPV
822	94-80-4	2,4-D-1 -butyl ester*	-1.79		4.38**	0.887	0.980	0	-0.495		N
823	789-02-6	o,p'-DDT*	-0.33		6.79**	1.150	1.019	0	1.086		N
824	67564-91-4	Fenpropimorph*	-1.30		4.93	1.180	0.907	0	2.798		HPV
825	319-84-6	HCH-alpha*	-2.33		3.80	0.000	1.085	0	-0.979		N
826	319-85-7	HCH-delta*	-2.32		3.78	0.000	1.095	0	-1.323		N
827	103055-07-8	Lufenuron*	-0.34		5.12	1.090	1.022	0	4.542		N
828	119168-77-3	Tebufenpyrad*	-1.61		4.61	0.882	0.950	0	1.683		N
829	16484-77-8	Mecoprop-P*	-2.74		2.94**	0.499	0.963	1	-1.186		HPV
830	1746-81-2	Monolinuron*	-1.82		2.30	1.282	0.978	0	0.305		N
831	52888-80-9	Prosulfocarb*	-1.82		4.65	0.830	0.929	0	1.501		N
832	52315-07-8	Zeta-cypermethrin*	-1.77		6.60	0.968	0.876	0	-0.520		N
833	66841-25-6	Tralomethrin*, §	-1.90	-2.48	7.56**	0.695	0.888	0	-1.669	-2.16	N
834	563-12-2	Ethion*	-1.22		5.07	1.086	0.971	0	0.934		N
835	70124-77-5	Flucythrinate*, §	-1.90	-2.71	6.20	0.997	0.864	0	-0.332	-2.43	N
836	52918-63-5	Deltamethrin*, §	-1.72	-1.10	6.20	0.771	0.921	0	0.071	-0.56	N
837	139968-49-3	Metaflumizone*	0.04		7.72**	0.819	0.944	3	5.518		N
838	70630-17-0	Metalaxyl-M*, §	-3.47	-2.62	1.71	0.718	0.906	0	-2.237	-2.32	N
839	108-62-3	Metaldehyde*	-2.44		0.12	1.229	0.946	0	2.076		HPV
840	422556-08-9	Pyroxsulam*, §	-2.46	-1.75	1.94	0.859	0.923	0	2.347	-1.31	N

Table E.1. Continued.

ID	CAS	Name	Pred pLC_{50} (QSTR)	Pred pLC_{50} (QTTR)	log K_{ow}	GATS7p	SpMaxA_ G/D	CATS2D_08_ DL	Mor31s	pEC_{50} (*D. magna*)[a]	HPV Status[b]
841	87820-88-0	Tralkoxydim*	-1.01		4.46	0.915	0.949	4	2.815		N
842	43121-43-3	Triadimefon*, §	-1.81	-0.88	2.77	1.216	0.947	0	1.349	-0.30	HPV
843	55219-65-3	Triadimenol*, §	-2.02	-1.58	2.90	1.189	0.936	1	-0.369	-1.12	HPV
844	2303-17-5	Triallate*, §	-1.84	-0.32	4.60	0.498	0.973	0	2.053	0.35	HPV
845	52-68-6	Trichlorphon (Chlorphos)*	-3.47		0.51	0.000	1.037	0	-0.335		N
846	80844-07-1	Etofenprox*, §	-1.70	-1.45	7.05	1.041	0.831	0	0.548	-0.97	N
847	102851-06-9	Tau-fluvalinate*	-1.12		6.81**	1.118	0.849	1	2.786		N

*QSTR external set chemical. §QTTR external set chemical. **Predicted log K_{ow} (EPI WSKOW v1.42). ***From this compound to the end: Chemicals with no ecotoxicological data (SU0303, 2015). [a]pEC_{50} (*D. magna*) data from Aalizadeh et al., 2017. [b]Production volume status according to OECD (2009). HPV: High production volume. N: Not HPV.