AN *IN SILICO* APPROACH FOR ESTIMATING THE ACTIVITY OF VECTOR CONTROL CHEMICALS TARGETING *Aedes aegypti* AND THEIR AQUATIC TOXICITY

by

Zeynep Yılmaz

Integrated B.S. and M.S. Program in Teaching Chemistry, Boğaziçi University, 2016

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

Boğaziçi University
2019

ACKNOWLEDGEMENTS

I am grateful for my family, my friends and my advisor Prof. Dr. Melek Türker Saçan for being such beauties in my life. I should have been blessed to have these people as my supporters. I would like to thank to Prof. Dr. Melek Türker Saçan for her endless support during my thesis study, she is the one that believes me and pushes me forward when I am lost.

I would like to thank to my parents İlhami Yılmaz and Şengül Yılmaz for their never-ending love and faith that I can feel whenever I need. I would like to thank to my brother Sinan Yılmaz for his psychological and financial support during my graduate years. I would like to thank to my sister Dilan Yılmaz for being such a lovely trouble in my life, my life would be meaningless without you.

I would like to thank to my best friend Burcu Calda for never leaving me alone, your friendship is the most valuable thing in my life. I would like to thank to my cousin Ezgi Marangoz for her great smile.

I would like to thank to my friends Ezgi Sunar, Selen Gökçe, Aygün Karaçay and Kamil Çöllü for their support during my thesis process.

ABSTRACT

AN *IN SILICO* APPROACH FOR ESTIMATING THE ACTIVITY OF VECTOR CONTROL CHEMICALS TARGETING *Aedes aegypti* AND THEIR AQUATIC TOXICITY

The mosquito *Aedes aegypti* is known as the main vector that transmits the viruses cause dengue, yellow fever, chikungunya epidemic arthritis, and Zika. Control of the vector is an important strategy to avoid disease propagation. However, vector control is threatened by the increasing resistance of mosquitoes to insecticides. On the other hand, environmental impacts of the intense use of these insecticides is of great concern. In the present study, the larvicidal activity of plant-derived compounds was subjected to a quantitative structure-activity relationship (QSAR) analysis. A valid QSAR model which fulfill the criteria set by the Organization for Economic Co-operation and Development (OECD) was generated using QSARINS 2.2.2 software. The generated QSAR model was validated both internally and externally. The external predictivity of model was tested with chemicals with no experimental larvicidal data and it has 95.3% structural coverage. The most toxic and the least toxic plant-based larvicides were determined. Piperidine derivatives were found highly effective on *Aedes aegypti* larvae. Also, the fruit *Piper nigrum* was highlighted as a plant-based larvicide source. Additionally, in order to propose a safe larvicide the toxicity of larvicides to non-target organism living in aquatic systems was evaluated by using previously generated acute toxicity and cytotoxicity models towards three representative aquatic species (algae, fish, and planarian) by Institute of Environmental Sciences, Ecotoxicology and Chemometrics Lab group and the most toxic larvicides are detected for these aquatic species.

ÖZET

*Aedes aegypti'*yi HEDEFLEYEN VEKTÖR KONTROL KİMYASALLARIN AKTİVİTESİNİN ve AKUATİK TOKSİSİTELERİNİN TAHMİNİ ÜZERİNE BİR *İN SİLİKO* YAKLAŞIM

Sivrisinek *Aedes aegypti* Zika, dengue ateşi, sarı humma, chikungunya artritine neden olan virüsleri ileten ana vektör olarak bilinmektedir. Vektörün kontrolü bu tür salgınların yayılmasını önlemek için önemli bir stratejidir. Bununla birlikte, vektör kontrolü, sivrisineklerin böcek öldürücülere karşı artan direnci nedeniyle tehdit altındadır. Öte yandan, bu böcek öldürücülerin yoğun kullanımının çevresel etkileri endişe vericidir. Bu çalışmada, bitki kaynaklı bileşiklerin larvisidal aktivitesi, kantitatif bir yapı-aktivite ilişkisi (QSAR) analizine tabi tutulmuştur. Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından belirlenen kriterlerine uygun geçerli bir QSAR modeli, QSARINS 2.2.2 yazılımı kullanılarak üretilmiş, üretilen model hem dahili hem de harici olarak doğrulanmıştır. Modelin harici tahmin özelliği deneysel veri içermeyen kimyasallarla test edilmiş, yapısal kapsamının %95.3 olduğu görülmüştür. En toksik ve en az toksik bitki bazlı larvisidler belirlenmiştir. Piperidin türevleri *Aedes aegypti* larvası üzerinde oldukça etkili bulunmuştur. Ayrıca, *Piper nigrum* bitkisi bitki bazlı larvisid kaynağı olarak dikkat çekici bulunmuştur. İlave olarak, güvenli bir larvisid önermek için larvisidlerin sucul sistemlerde yaşayan hedef olmayan organizmalara toksisitesi, daha önce Çevre Bilimleri Enstitüsü, Ekotoksikoloji ve Kemometri Laboratuvarı grubu tarafından geliştirilen üç temsili sucul türe (alg, balık ve planaryan) yönelik akut toksisite ve sitotoksisite modelleri kullanılarak değerlendirilmiş ve üç organizma çeşidi için en toksik larvisidler belirlenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation	Unit
E_{aq}	Energy in Aqueous Phase	eV
EC_{50}	Concentration of a Chemical that Causes %50 Effect on a Target Organism	mM
E_{HOMO}	The Highest Occupied Molecular Orbital Energy	eV
E_{LUMO}	The Lowest Unoccupied Molecular Orbital Energy	eV
HATS1s	Leverage-weighted Autocorrelation of lag 1 / Weighted by I-state	
H-046	H aAttached to C0(sp3) no X Attached to Next C	
*h**	Critical Hat Value	
F	Fischer Statistics	
K_{ow}	*n*-Octanol/Water Partition Coefficient	
LC_{50}	Concentration of a Chemical that Causes %50 Lethality a Target Organism	M
Log K_{ow}	Logarithm of K_{ow}	
Mor30s	Signal 30 / weighted by I-state	
Q^2_{LOO}	Leave-One-Out Cross Validation	
Q^2_{LMO}	Leave-Many-Out Cross Validation	
R^2	Coefficient of Determination	
R^2_{adj}	Adjusted (for degrees of freedom) R^2	
UI	Unsaturation Index	

Abbreviation	**Explanation**
AD	Applicability Domain
CAS	Chemical Abstracts Service
CCC	Concordance Correlation Coefficient
EC	European Commission
ED	Euclidean Distance
GA	Genetic Algorithm
GATS7e	Geary Autocorrelation of lag 7 Weighted by Sanderson Electronegativity
GETAWAY	Geometry, Topology, and Atom-Weights Assembly
MAE	Mean Absolute Error
MCDM	Multi-Criteria Decision Making
MLR	Multiple Linear Regression
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
PM6	Parameterized Model 6
QSAR	Quantitative Structure-Activity Relationship
QUIK	Q^2 Under Influence of K
REACH	European Regulation for the Registration, Evaluation, Authorization and Restriction of Chemicals
RMSE	Root Mean Square Error
RTL-W1	Rainbow Trout Liver Cell Line
US EPA	United States Environmetal Protection Agency
3D-MoRSE	3D-Molecule Representation of Structures based on Electron Diffraction

1. INTRODUCTION

In today's world vector borne diseases are an emerging issue considering the fact that causing more than 700 000 deaths annually. According to World Health Organization (WHO), more than 3.9 billion people in over 128 countries are faced with the risk of dengue, with 96 million cases estimated per year (WHO, 2017). *Aedes aegypti* a blood-sucking insect that lives in tropical and subtropical areas is the principal vector for dengue virus transmission. This species is also considered to transmit chikungunya, yellow fever and Zika virus. The risk of outbreaks is growing so fast with climate change. According to the WHO, it is estimated that this mosquito species causes 50 million infections and 25 000 deaths per year (WHO, 2018).

To control the virus transmission by *Aedes aegypti*, one of the most common methods is to use larvicides, but the intense use of larvicides may result with both environmental contamination and resistance of the mosquitoes to these chemicals. For example; one of the most employed larvicide temephos is resulted with the resistance-gained by mosquito (Melo-Santos et al., 2010). It is also stated that organophosphorus compounds like temephos has an activity in non-targeted organisms too (Saavedra et al., 2018). So, the need for the new larvicides without such side effects is an emerging issue. Many publications have recently reported new larvicides as both synthetic and plant derived (Hansch & Verma, 2009; Pohlit et al., 2011; Santos et al., 2010). Among these, the compound which prevents insect resistance and environmentally friendly will be remarkable and useful. However, finding this larvicide requires large budget, time and human effort. Therefore, to reduce these requirements quantitative structure activity relationship (QSAR) is a way for predicting the properties and/or biological activities of chemicals not synthesized yet. The studies about the quantitative structure-activity relationships (QSAR) have been gaining importance to save time and cost involved in identifying candidate vector control chemicals with larvicidal activity (Devillers et al., 2014 & 2015).

Even if there are not many QSAR studies on larvicidal activity of plant derived compounds, recently published QSAR studies about *Aedes aegypti* larvicides cannot be considered as valid because of having narrow range of endpoint values (LC_{50}) (Carmenate et al., 2017; Doucet et al., 2017; Saavedra et al., 2018). It is well known that an endpoint range of at least 2.0 log unit is required to consider the dataset for generation of a QSAR model (Cronin et al., 2009). Additionally, the generated literature larvicidal QSAR models were not validated by using up-to-date validation criteria

reported in the literature. Also, even if the application area of a larvicide is the breeding areas like standing water, shallow ponds, lakes, woodland pools, marshes and swamps, previous studies generally don't contain any aquatic toxicity data of vector control chemicals.

1.1. Aim of the study

One of the purposes of the present study is to develop an *in silico* model particularly a QSAR model using the larvicidal activity values of various plant-based chemicals compiled from the literature Since the dataset range of the literature QSAR model doesn't meet the need for at least 2.0 log difference for activity range, we aim to carry out an extensive research on the activities of plant-based larvicides and generate a reliable QSAR model. The main steps to reach the purpose of this study are the generation of a QSAR model according to the following scheme:

1. to split the compiled dataset into training/test sets for the generation of a linear QSAR model;
2. to calculate the theoretical molecular descriptors representing the molecular structures in the dataset using DRAGON 7.0 (Talete Inc., 2017) and SPARTAN 16 (Wavefunction, 2016) software packages
3. to select descriptors from the large descriptor pool using "All subset", "genetic algorithm (GA)" and "Hold model and add one more variable" tools of QSARINS software (v.2.2.2) (Gramatica et al. 2013, 2014; QSARINS 2017);
4. to validate the model internally and externally (using the test set and up-to date validation metrics)
5. to define applicability domain of the generated QSAR model using the leverage approach by highlighting both the response-outliers and the structural influential chemicals (Williams graph).
6. to predict the larvicidal activity values of various plant-based chemicals with no reported data in the literature.

Another aim of this study to predict the aquatic toxicity of plant-based vector control chemicals/ larvicides with no experimental data that compiled from the literature. For this purpose, acute toxicity and cytotoxicity models towards three aquatic species (algae, fish and planarian) were used. Relevant to this part, we aim to screen the most and the least toxic larvicides against these three species using their structure-activity relationships. Lastly, if possible, another purpose of this study is to propose both effective and environmentally safe larvicide.

2. THEORETICAL BACKGROUND

2.1. *Aedes aegypti* mosquito control

Vector borne diseases equal for more than 17% of all infectious diseases, more than 700 000 deaths annually are caused by vector borne diseases. One of the most important disease caused by vectors is dengue. Dengue virus is transmitted mainly by female mosquitoes of *Aedes aegypti*. Dengue is found at tropical and sub-tropical climates worldwide, and it is stated that the global incidence of dengue has been gradually increasing recent decades (WHO, 2019). Also, it is stated that the half of the world's population is now at risk of dengue even if there is no specific treatment for dengue (WHO, 2019). *Aedes aegypti* is also responsible for serious illnesses like chikungunya, lymphatic filariasis, yellow fever and Zika (WHO, 2017). Zika virus infection is associated with an increased risk of neurologic complications in adults and children, especially for infants the Zika virus infection during pregnancy can cause microcephaly and other congenital malformations (WHO, 2018). Since there is no vaccine or drug currently available to treat both dengue and Zika virus' infections, effective vector control measures have gained importance recent years.

Aedes aegypti is a vector that adapted to urban environments and human habitation areas are the main locations for these species. *Aedes aegypti* is a holometabolous insect which means that the insect has a metamorphosis that goes through an egg, larvae, pupae, and adult stage. The life span of the insect depends on the environmental conditions, mostly on temperature. The larval stage consisted of four stages called instars. Larvae spends short amount of time in the first three instar compared to the last instar. Although the larval stage of *Aedes aegypti* is mostly spent on the water surface, while feeding larvae moves on the bottom of the container to reach organic particulate matter such as algae, other microscopic organisms (Nelson, 1986).

The effective control mechanism for vector borne diseases starts with the disrupting biological cycle of vectors. (Goellner et al., 2017). The ways that can break this biological cycle are the main topic of current studies. Larvicides, biological agents and insecticides have been studied lately for vector control. However, biological agents are not demanded on the market due to both the high cost of production and the resistance gained by mosquito strains (Federici et al., 2003 & Paris et al., 2011).

On the other hand, the excess use of chemical agents may result not only in danger for human health and aquatic life but also in resistance vectors, too. Temephos is the most used larvicide so far,

but the intense use of this compound caused both mosquito-resistance and aquatic toxicity. Resistance to temephos has been seen a lot of countries of the world (Pandey et al., 2013). Aquatic toxicity of temephos has been studied by Abe and colleagues (2014) and the study results that temephos represent high toxicity against *Daphnia magna* with 48h EC_{50} = 0.15 μg/L and this larvicide shows high environmental risk to this species (Abe et al., 2014). One of the most used adulticide malathion has also proven with the result of insecticide resistance in different studies (Hidayati et al., 2011, Goindin et al., 2017). In addition to this, malathion (Figure 2.1) is restricted from European market due to the monograph of WHO International Agency for Research on Cancer that states malathion as carcinogenic for humans (Guyton et al, 2015).

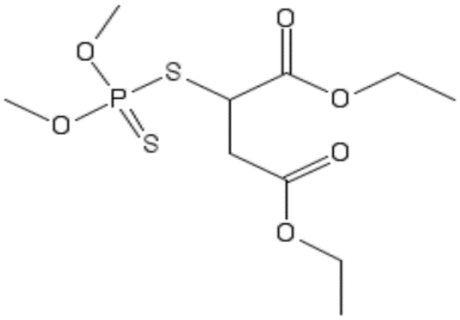

Figure 2.1. Chemical structure of malathion (Structure was drawn using PubChem Sketcher v.2.4).

These reasons above show that precautions for the vector borne diseases should be taken seriously regarding human health, environmental effects and economic issues. Therefore, nowadays finding alternative methods for controlling vectors has become an urgent issue. Searching environmentally safe, low cost and potentially high chemical agents for mosquito controlling makes plants a current topic. Today plants have been investigated as a source for alternative agents to control of mosquitoes because of containing bioactive compounds and being eco-friendly (Shivakumar et al., 2013).

2.2. Plant-based larvicides

Plants luckily have a rich source of bioactive compounds, so extracts and/or essential oils from plants might contain alternative compounds for vector controlling. Also plant extracts and/or essential oils are environmentally friendly because they are easily biodegradable into nontoxic compounds (Liu et al., 2006). Besides, for thousands of years with the co-evolution of plants and insects, plants gained both chemical and physical mechanisms to protect themselves against insects. That's why different plant parts (leaves, root and bark) have been used by humans to control vectors (Dias et al.,

2014). For example, United States Environmental Protection Agency (US EPA) stated that citronella oil that is an essential oil extracted from the roots and stems of the Cymbopogon (lemongrass) as a nontoxic insect repellent (EPA, 1999). All of the reasons make plants are extremely worth to investigate (Dias et al., 2014).

Plant extracts involve substances with insecticidal activity, these substances are called as essential oils (Pavela, 2015). Essential oils that include terpenes and phenylpropanoids are natural combinations of volatile organic compounds. They are considered to be one of the best ways for vector controlling naturally (Kweka et al., 2016). A number of studies has been conducted in order to find a larvicide in natural products mainly in essential oil of plants (Doria et al., 2009; Govindarajan et al., 2010; Lucia et al., 2007; Perumalsamy et al., 2009; Santos et al., 2010; Silva et al., 2008).

Not only essential oils from plants but also piperine alkaloids and piperidine derivatives found in *Piper* species have been studied for larvicidal, insecticidal and repellant activity (Park et al., 2002). Studies indicate that *Piper* species have bioactive metabolites including alkaloids, flavonoids, amides and terpenoids. These metabolites have been investigated further because of their therapeutic and commercial value (Marques and Kaplan, 2015).

2.2.1. Monoterpenes and sesquiterpenes

Terpenes are the largest single class of natural compounds with a wide range of biological activities found in essential oils. Terpenes are made from isoprene molecules which is consisted of five carbon atoms with double bonds $(C_5H_8)n$. The simplest terpenes, monoterpenes, contain two isoprene units; and sesquiterpenes contain three isoprene units. Terpenes are mostly hydrocarbons, but they also include oxidation products like alcohol, ketones and aldehydes.

Monoterpenes and sesquiterpenes have a great activity as deterrent for herbivores (Chizzola, 2013). Santos and colleagues (2011) investigated the larvicidal activities of 14 monoterpenes and oxygenated monoterpenes against 3rd instar Aedes aegypti larvae. The study showed that both *R*-limonene (Figure 2.2.(a)) and *S*-limonene (Figure 2.2.(b)) have high potency as larvicide with the 24h- LC_{50} values of 27 ppm and 30 ppm, respectively. The least potent compound was found as menthone with LC_{50} value of 508 ppm.

(a) (b)

Figure 2.2. Chemical structure of (a) *R*-limonene and (b) *S*-limonene as monoterpenes (Structures were drawn using PubChem Sketcher v.2.4).

Another study about of plant-based larvicides was done by Perumalsamy and colleagues (2009) using early 3rd instar larvae of *Aedes aegypti* with the extracted compounds isolated from *Asarum heterotropoides*. The study indicated that safrole (9.88 ppm) is the most potent larvicide against *Aedes aegypti* larvae followed by two monoterpene hydrocarbons (-) -(β)- pinene and γ-terpinene with 24h-LC_{50} values of 15.40 and 17.11 ppm, respectively.

2.2.2. Phenylpropanoids

Monoterpenes and sesquiterpenes are mainly the major constituents of essential oils, but in some plant species phenylpropanoids are found too, sometimes they are as the main component (Freidrich, 1976). Phenylpropanoids provide plant responses like stress upon variation of light and plant resistance towards herbivores (Vogt, 2010).

Hematpoor and colleagues (2016) investigated the larvicidal, ovicidal and AChE inhibition effects of three phenylpropanoids against late 3rd or early 4th instar larvae of three vector species *Aedes aegypti*, *Aedes albopictus* and *Culex quinquefasciatus*. The studied compounds were asaricin, isoasarone and trans-asarone. Asaricin and isoasarone were found as highly potent against three larvae species with ≤ 15 μg/mL for 100% mortality. Ovicidal activity of the compounds were evaluated via egg hatching, both asaricin and isoasarone showed ovicidal activity. The study also provided the results that proves these two compounds as neuron toxic toward three species with high AChE inhibiton IC_{50} values of 0.73 to 1.87 μg/mL, respectively.

Another study about phenylpropanoids against *Aedes aegypti* was carried by Pinto and colleagues (2012), the adulticidal activity of dillapiole (Figure 2.3) and its derivatives were investigated. The study results showed that dillapiole and isodillapiole were the most potent compounds (100% mortality after exposure for 45 min) to *Aedes aegypti* adult females with 90 minutes exposure whereas dillapiole derivatives showed lower adulticide potency.

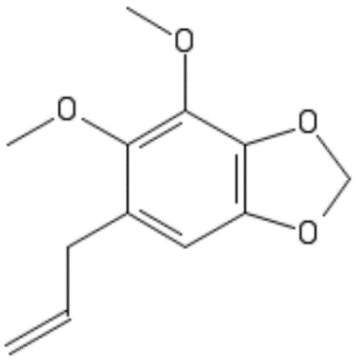

Figure 2.3. Chemical structure dillapiole (Structure was drawn using PubChem Sketcher v.2.4).

2.2.3. Alkaloids

Alkaloids are thought to be the largest class of compounds that plants produce, and they are the part of defense mechanism of the plants against herbivores. The structure of the alkaloids mainly involves one or more nitrogen atom, whereas piperidine alkaloids contain the piperidine nucleus. Piperidine (Figure 2.4) is a naturally occurring compound mainly found in *Piper nigrum* L., *Piperaceae* plants and piperidine derivatives are subjected to various studies due to their warding off effect against herbivores (Ojima, 1999). Park and colleagues (2002), examined the larvicidal activity of four compounds derived from the fruits of *Piper nigrum* against third instar larvae of *Aedes aegypti* for 48h duration. They reported that the larvicidal activity against *Aedes aegypti* larvae, was more pronounced by retrofractamide A (0.039 ppm) than pipercide (0.1 ppm), guineensine (0.89 ppm), and pellitorine (0.92 ppm).

Also, Pridgeon and colleagues (2007), investigated the insecticidal activities of 33 derivatives of piperidine against female *Aedes aegypti* adults and the structure activity relationship of these compounds. The study showed that different moieties on the piperidine ring cause different adulticide effect against *Aedes aegypti*.

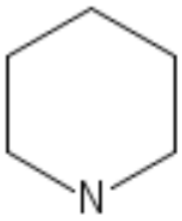

Figure 2.4. Chemical structure of piperidine (Structure was drawn using PubChem Sketcher v.2.4).

2.3. Aquatic toxicity of vector control agents

Aquatic toxicity testing is used to determine if a compound pose a risk to aquatic environment. Because aquatic toxicity data can also be used as a base for other environmental areas like soil and sediment, information about aquatic toxicity of chemicals has been gained importance in regulatory purposes. European Commission has introduced a legislation about the use and impact of chemicals on both human health and environment (EC, 2006). So, the need for investigation of the toxicity of chemicals in order to protect the aquatic environment and to find environmentally safe larvicide makes scientists to do toxicity testing experiments.

Preventing the development of immature larva from becoming an adult mosquito makes larvicides the most important strategy for mosquito control (Nunes et al., 2018). *Aedes aegypti* mosquitoes can reproduce easily in every area of standing water collection so the application area of the larvicides are these places but the accessibility of chemical agents through water streams into aquatic environment is a great concern. Although there were not so much studies about the aquatic toxicity of larvicides against nontarget organisms, the need for environmentally safe larvicide makes the situation worth to investigate.

The aquatic toxicity of biologic agents was investigated in different studies but there was negligible environmental impact found. The study carried by Lagadic and colleagues (2014), showed that even if the repeated use of Bti over many years creates questions about the possible long-term effects on nontarget organisms, the long-term use of Vectobac® had almost no impact on nontarget aquatic invertebrates compared to other abiotic factors. Aquatic toxicity of biological agents on nontarget wetland invertebrates has also been studied by Merritt and colleagues (2005) and this 3-year study evaluated the aquatic toxicity in 5 different metrics: mean taxa richness, mean diversity, Diptera richness, Diptera abundance, functional group changes in percent. The study has also revealed

that there were no detrimental effects created by *Bacillus sphaericus* (VECTOLEX®) on nontarget organisms.

There are also studies that reveals the aquatic toxicity of plant-based larvicides against nontarget organisms. The aquatic toxicity of *Artemisia absinthium* essential oil (EO) and its three major chemical constituents was evaluated and the results showed that both EO and its major constituents (E)-β-farnesene, (Z)-en-yn-dicycloether, and (Z)-β-ocimene have moderate toxic effect on-target organisms *Chironomous circumdatus, Anisops bouvieri* and *Gambusia affinis* (Govindarajan and Benelli, 2016). Another study carried by Pavela and Govindarajan (2016) investigated the aquatic toxicity of *Zanthoxylum monophyllum* leaf essential oil (EO) and its major chemical constituents against non-target fish *Gambusia affinis*. The study showed that the EO and its major constituents Germacrene D-4-ol and α-Cadinol were found safer to *Gambusia affinis* (Pavela and Govindarajan, 2016).

The studies about aquatic toxicity of vector control chemicals are not enough to find environmentally safe larvicide, considering the number of chemicals both plant-based and synthetic. It is hard to do *in vivo* and *in vitro* studies regarding the money, time and source need. That's why *in silico* approach has gained importance as an alternative for laboratory testing. Registration, Evaluation and Authorization of Chemicals (REACH) has been making attempts to reduce the number of animals tested for risk assessment of chemicals. Therefore, quantitative structure activity relationships (QSARs) is a way for predicting the toxicity of chemicals to meet the need for data in ecotoxicity.

2.4. QSAR modelling

QSAR studies are based on the idea of finding associations between chemical structures and biological activity (Veerasamy, 2011). QSAR models are applicable for assessing the potential effects of compounds on human health and environment without doing laboratory experiments to meet the need for a wide range of chemicals without toxicological and ecotoxicological data. Also, QSAR studies are helpful in case of screening, prioritization and identification of chemicals which are required by REACH in order to manage the production of chemicals that might pose threat to human health and environment.

Basically, a QSAR model shows a mathematical equation that correlates the response of chemicals with their structural information in the form of numbers, i.e. "molecular descriptor". This

correlation depended on the structural information that derived by 2D and/or 3D molecular properties, geometric, topological, chemical properties etc. The steps for the generation of a QSAR model is shown in Figure 2.5.

QSAR studies have been used for 50 years since the study about pesticides done by Hansch and colleagues (1962). In these 50 years, many QSAR models developed and there have been many statistical methods and validation techniques generated to originate a robust mathematical equation. In the present study Multiple Linear Regression (MLR) based on ordinary least squares (OLS) method was used. MLR is a commonly used method used in QSAR that uses different variables to predict the response of a variable with a linear equation. An MLR equation can be like the following (Eq 2.1):

$$Y = a_0 + a_1 \times X_1 + a_2 \times X_2 + \cdots + a_n \times X_n \quad (2.1)$$

where Y is the response variable, X_1 , $X_2 \ldots X_3$ are descriptors (independent variables) and a_1, $a_2 \ldots a_n$ are the regression coefficients and finally a_0 is the constant term of the model.

Considering the plenty of descriptors calculated, the use of all numerous combinations of the available ones for model calculation by means of the MLR would be impossible. In the present study, all the possible combinations of the selected descriptors are explored via the "All Subset" method in QSARINS. The best linearly correlated combinations are listed by the software in terms of leave–one–out cross–validated R^2 (Q^2_{LOO}). Genetic Algorithm (GA) is one of the most preferable methods for the selection of descriptors, because of its superior performance in variable selection. It is an adaptive heuristic search algorithm based on evolutionary ideas of natural selection and genetics. It combines survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. GA defined as a search approach which uses random choice as a tool to guide a highly exploitative search through a coding of a parameter space (Goldberg, 1988). The mechanistic of a simple genetic algorithm is involving copying strings and swapping partial strings based on three operators:

- Reproduction
- Crossover
- Mutation.

QSARINS employs Tournament Selection method to select best representative descriptors via GA.

In order to develop a valid QSAR model, the validation guidelines set by OECD to interpret a QSAR model as reliable and acceptable are (OECD, 2007):

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

The first principle is about "endpoint" that is the measure of target activity. In order to have a reliable QSAR model, given endpoints should be consistent in a dataset. That's why it is important collect data from the experiments with same protocols while generating a QSAR model (OECD, 2007). For a valid QSAR model the methodology should be clearly defined. In the present study the endpoint was stated as lethal concentration 50 (LC_{50}) which means the concentration required to kill half the members of a tested population after a specified test duration. According to World Health Organization (WHO), to evaluate the biological activity of a mosquito larvicide, laboratory-reared mosquito larvae of known age or instar (reference strains or F1 of field-collected mosquitoes) are exposed for 24 h to 48 h or longer in water treated with the larvicide at various concentrations within its activity range, and mortality is recorded. Probit analysis is used for the determination of the lethal concentration of the larvicide for 50% (WHO, 2005). As stated in the second principle, the stages like descriptor selection, training/test set divisions and statistical parameters should be clear in a valid QSAR model. The third principle mainly deals with the applicability domain which defines the limits of the model in the endpoint predictions. In the present study the main gap in the literature has been filled with a good range of endpoint values. In the fourth principle, the appropriate measures of goodness of fit, robustness (internal validation) and predictivity (external validation) which were further discussed in the following section, stated as requirements for a valid QSAR model. Finally, mechanistic interpretation is the explanation of the model via molecular descriptors in order to represent the reasons behind the activity of compounds.

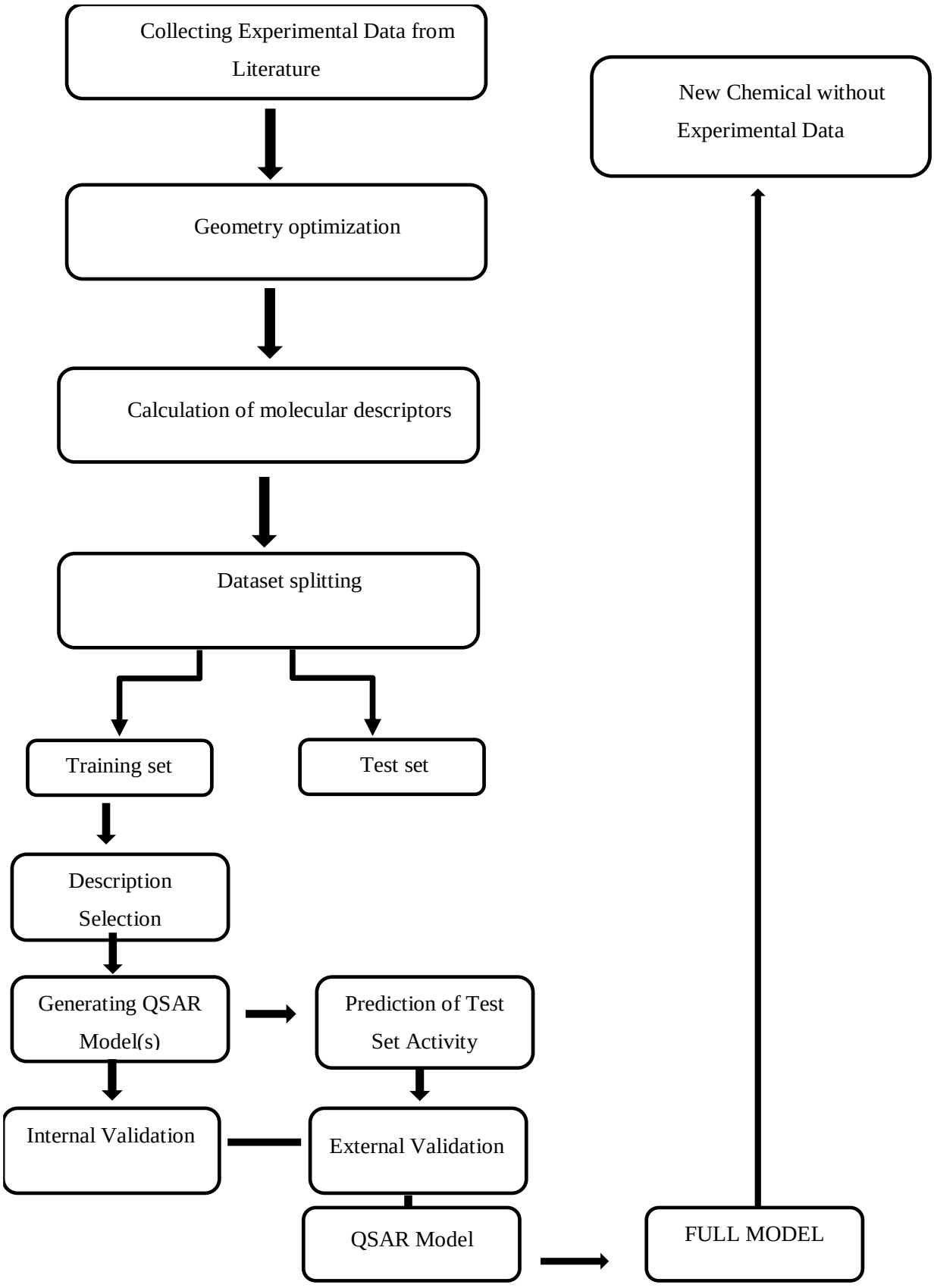

Figure 2.5. Basic steps of a QSAR model.

2.5. Internal validation parameters:

The statistical performances of the generated QSAR models were evaluated by the determination of coefficient (R^2), leave one-out cross-validation coefficient (Q^2_{LOO}), Fisher criterion (F), root mean square error of training set ($RMSE_{Tr}$) and Y-randomization test (shuffling of 2000 times). Highest R^2, greatest Q^2_{LOO} were preferred because these parameters measure the fitting and robustness of the a QSAR model, respectively.

2.5.1. Determination Coefficient (R^2) and Adjusted Determination Coefficient (R^2_{adj})

The determination coefficient R^2 can be defined in the following equation (Eq. 2.2):

$$R^2 = 1 - \frac{\Sigma (Y_{obs} - Y_{calc})^2}{\Sigma (Y_{obs} - \overline{Y_{obs}})^2} \quad (2.2)$$

where Y_{obs} is the observed response value, while Y_{calc} is the model-derived calculated response and $\bar{Y}_{obs}$ is the average of the observed response values. For an ideal model R^2=1 where the sum of squared residuals equals 0. The fitting quality of the model increases as the value of R^2 closes to 1, but if R^2 > 0.60 the model can be considered as acceptable (Golbraikh and Tropsha, 2002).

Increasing the number of descriptors for a model leads an increase in the value of R^2, but this increase also creates problem in statistical reliability and in the degree of freedom. To fix this problem, R^2_{adj} is used to show the fraction of the data variance explained by the model.

2.5.2. *F* (Variance Ratio)

The Fischer statistics value F which is used to decide the overall significance of the regression coefficients can be defined in the following equation (Eq. 2.3):

$$F = \frac{\frac{\Sigma (Y_{calc} - \bar{Y})^2}{p}}{\frac{\Sigma (Y_{obs} - Y_{calc})^2}{N-p-1}} \quad (2.3)$$

where N is the number of experimental response values and p is the number of independent variables used to predict the response values and $\bar{Y}$ is the average of the experimental response values. Higher F value shows more significant model.

2.5.3. *Y*-scrambling

Response randomization *(Y*-scrambling*)* is a technique that shows a possible chance correlation between independent variables and a response variable. *Y*-scrambling is performed in order to determine the robustness of a QSAR model. This is done by intentionally destroying the connection between target variable *Y* and independent variables *X* (molecular descriptors in QSAR) by randomly permuting the *Y* data, leaving all *X* data untouched, and performing the whole model building procedure as it would be done for real *Y* data (Rücker et al., 2015). This technique must be used in accordance with cross-validation (CV) and must always be applied to check the significance of the developed QSAR model obtained by chance correlation (Gramatica, 2007).

2.5.4. Multicollinearity between descriptors

The QUIK rule (Q Under Influence of K) has been demonstrated to be very effective in avoiding models with multicollinearity without prediction power (Todeschini et al., 1999). QUIK rule is a simple criterion based on the K multivariate index that allows the rejection of models with high predictive collinearity that can lead to chance correlation (Todeschini et al., 2004). This rule is derived from the evident assumption that the total correlation in the set given by the model predictors X plus the response Y (K_{XY}) should always be greater than that measured only in the set of predictors (K_X) (Todeschini et al., 2004). Therefore, the QUIK rule is: only models with the K_{XY} correlation among the [X + Y] variables greater than the K_X correlation among the [X] variables can be accepted, or if $[K_{XY}] - [K_X] < \delta K$ reject the model (Todeschini et al., 2004). δK (Delta K) is a limit defined by the user. It was set to 0.05 to minimize the inter–correlation among descriptors. Additionally, to eliminate chance correlations and unstable model, the ratio of number of training set compounds to the number of descriptors in a linear QSAR model should be at least 5:1. This criterion is called “Topliss and Costello rule” (Topliss and Costello, 1972).

2.5.5. Root mean squared error of training set ($RMSE_{Tr}$)

Root mean squared of error of training set ($RMSE_{Tr}$) shows the overall error of the model for the training set. $RMSE_{Tr}$ is performed to measure and compare the accuracy of the proposed QSARs.

2.5.6. Leave-one-out (LOO) cross-validation (Q^2_{LOO})

The cross-validation leave-one-out is used as a fitness function during the model development step. One compound is excluded from the training set and the response of the excluded compound is predicted by the model. In other words, it assesses the ability of the model to predict new chemicals in the data set one by one, putting them iteratively in the test set. The value greater than 0.5 is generally regarded as good. The addition of descriptors may be continuous till the increase in the number of descriptors does not efficiently improve the Q^2_{LOO} value. The formula of Q^2_{LOO} value is as follows (Eq. 2.4):

$$Q^2_{LOO} = 1 - \frac{\Sigma (Y_{obs} - Y_{pred})^2}{\Sigma (Y_{obs(training)} - \bar{Y}_{training})^2} \quad (2.4)$$

where Y_{obs} and Y_{pred} refer to observed and LOO-predicted activity values. $Y_{obs\,(Tr)}$ is the observed activity and $\bar{Y}_{Tr}$ is the average of experimental response values for the training set compounds.

2.6. External validation parameters:

The predictive abilities of the models were evaluated by external validation parameters. The reliabilities of the models were also judged by additional validation parameters known as Golbraikh and Tropsha's criteria (Golbraikh and Tropsha, 2002).

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

Q^2_{F1} value which is used to show the degree of correlation between observed and predicted activity data can be defined in the following equation (Eq. 2.5) (Shi et al., 2001):

$$Q^2_{F1} = 1 - \frac{\Sigma (Y_{obs(test)} - Y_{pred(test)})^2}{\Sigma (Y_{obs(test)} - \bar{Y}_{training})^2} \quad (2.5)$$

where $Y_{obs(test)}$ and $Y_{pred(test)}$ refer to observed and predicted activity values for test compounds. $\bar{Y}_{training}$ is the average of experimental response values for the training set compounds.

Q^2_{F2} which is based upon prediction of test set compounds is generated by (Schüürmann et al. 2008) and given in the following equation (Eq. 2.6):

$$Q^2_{F2} = 1 - \frac{\Sigma\,(Y_{obs(test)} - Y_{pred(test)})^2}{\Sigma\,(Y_{obs(test)} - \bar{Y}_{test})^2} \quad (2.6)$$

In the equation the $\bar{Y}_{test}$ is the mean experimental response for the test set compounds.

Q^2_{F3} which measures the model predictability is proposed by Consonni et al. (2009, 2010) and can be defined in the following equation (Eq. 2.7):

$$Q^2_{F3} = 1 - \frac{[\Sigma\,(Y_{obs(test)} - Y_{pred(test)})^2]/n_{test}}{[\Sigma\,(Y_{obs(training)} - \bar{Y}_{training})^2]/n_{Tr}} \quad (2.7)$$

where n_{test} and n_{Tr} refer to the number of compounds in the test and training set, respectively. A threshold value 0.7 is defined for these parameters above (Chirico and Gramatica, 2011).

2.6.2. Root mean squared of error of test set ($RMSE_{test}$)

$RMSE_{test}$ used for the external predictive ability of the proposed model is given in the following equation (Eq. 2.8):

$$RMSE_{test} = \sqrt{\frac{\Sigma\,(Y_{obs(test)} - Y_{pred(test)})^2}{n_{test}}} \quad (2.8)$$

2.6.3. Concordance Correlation Coefficient (*CCC*)

Concordance Correlation Coefficient (*CCC*) parameter which is calculated both for training and test set data in order to check the reliability of the model can be defined in the following equation (Eq, 2.9) (Lin 1989, 1992):

$$CCC_{test} = \frac{2\Sigma\,(Y_{obs(test)} - \bar{Y}_{obs(test)})(Y_{pred(test)} - \bar{Y}_{pred(test)})}{\sum_{i=1}^{n}(Y_{obs(test)} - \bar{Y}_{obs(test)})^2 + \sum_{i=1}^{n}(Y_{pred(test)} - \bar{Y}_{pred(test)})^2 + n(Y_{obs(test)} - \bar{Y}_{pred(test)})} \quad (2.9)$$

where $Y_{obs(test)}$ and $Y_{pred(test)}$ refer to experimental and predicted activity values for the test compounds, n is the number of compounds, $\bar{Y}_{obs(test)}$ and $\bar{Y}_{pred(test)}$ refer to averages of observed and predicted values of the test compounds, respectively. CCC should be 1 for the ideal model, but the threshold value is set as 0.85 (Chirico and Gramatica, 2012).

2.6.4. r^2_m Metrics

The r^2_m and Δr^2_m parameters are used to assess the performance of the models and these parameters can be defined in the following equations (Eq. 2.10 & 2.11):

$$r_m^2 = r^2(1 - \sqrt{r^2 - r_0^2}), \tag{2.10}$$

$$\Delta r_m^2 = |r_m^2 - r_m'^2| \tag{2.11}$$

where r^2 is the determination coefficient for the test set with an intercept and r_0^2 is the determination coefficient without an intercept, $r_m'^2$ is the determination coefficient for the experimental activity value on the x-axis and predicted activity value on the y-axis. The threshold values for these two parameters are $r_m^2 > 0.50$ and $\Delta r_m^2 < 0.20$ (Ojha et al., 2011).

2.6.5. Mean Absolute Error (*MAE*)-Based Criteria

MAE is the mean absolute error is another external parameter used in order to check the errors in predictions of the generated models. *MAE* can be calculated by the following equation (Eq. 2.12):

$$MAE = \frac{1}{n} x \left| Y_{obs} - Y_{pred} \right| \tag{2.12}$$

where Y_{obs} and Y_{pred} refer to observed and predicted activity values and n is the number of compounds. MAE_{test} is the mean absolute error which is calculated for 95% of the test set data when $n_{test} > 10$ (Roy et al., 2016).

Regarding the *MAE*-based criteria, an ideal model should fulfill the following parameters:

$MAE_{test} \leq 0.1 \times$ training set range (TSR)

$MAE_{test} + 3 \times \sigma \leq 0.2 \times$ TSR

where σ is the standard deviation of the absolute error values of the test data.

2.6.6. Golbraikh and Tropsha Criteria

According to Golbraikh and Tropsha (2002), if the following criteria are satisfied, the model can be considered as acceptable:

a) $Q^2_{Tr}>0.5$

b) $R^2>0.6$

c) $(R^2 - R^2) / R^2 < 0.1$ 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

d) $|R_0^2 - R_0'^2|<0.3$

where R_0^2 (predicted vs. observed) and $R_0'^2$ (observed vs. predicted) are the determination coefficients without intercept, *k* and *k'* are the slopes.

2.7. Literature QSAR Models on Larvicidal Activity

There are several studies in order to derive QSAR models for vector control chemicals with larvicidal activity. For example, Devillers and colleagues (2014) conducted a study about all the existing in silico models for predicting vector control chemicals targeting *Aedes aegypti* up to the date. In this paper, the existing models about larvicides and adulticides like juvenile hormone mimics, organotin compounds, ecdysteroids were evaluated according to their differences in strategy to develop a new larvicide or adulticide. In 2015, Devillers and colleagues studied the structure–activity relationship (SAR) modelling of juvenile hormone activity of *Aedes aegypti* against structurally diverse chemicals. They used the dataset consisted of 188 chemicals with their activity against *Aedes aegypti* larvae as IC_{50} (concentration required to produce 50% inhibition of larval development, mmol). At the end of the study, from the different modelling results they propose new chemicals for synthesis.

Doucet and colleagues (2017) developed QSAR models for larvicidal activity prediction of piperidine derivatives against *Aedes aegypti*, but as it stated in the article the dataset of the study was to narrow that the range for pLD_{50} values was 1.01-2.48 (less than 2 log unit). Also, the diversity of the structures of the chemicals was too sparse for developing a good QSAR model. Despite these constraints, they reported a QSAR model with R^2 of 0.860 by using 2D topology-based descriptors calculated with PaDEL software.

A mathematical model for larvicidal activity prediction of 55 compounds was developed by Carmenate and colleagues (2017) using QSARINS software and this model has an R^2 of 0.752 which is relatively high compared to other models in the literature. However, this model has some drawbacks. One of the drawbacks is the min. and max. values for pLC_{50} (pLC_{50} = -log (LC_{50}) used in this study which are 2.04 and 3.85 (mol/L), respectively, (less than 2.0 log unit) indicating a narrow range for endpoint values in terms of QSAR modelling principles. Filho and colleagues (2016) developed a QSAR model for larvicidal activity of 31 monoterpenes and structurally related compounds which are bioactive against *Aedes aegypti*. Although the model has a high R^2 as 0.830 together with other statistical parameters, the distribution of test and training set chemicals around the fitted line wasn't good.

Another issue is the unit of lethal concentration, some studies show LC_{50} value in ppm rather than in molar unit which makes difficult to compare their larvicidal activity. The current study about QSAR analysis of 60 plant derived compounds was done by Saavedra and colleagues (2018) using freely available descriptors. Even if the model has an R^2 of 0.84, both the activity range which is less than 2.0 log unit and the unit of pLC_{50} value (μg/mL) make this study not easily applicable regarding with the requirements for a valid QSAR model.

3. MATERIALS AND METHODS

3.1. Dataset:

Dataset compiled from the literature on larval toxicity of plant-based compounds are listed in Table 3.1 and used for QSAR modeling. In the present study, the endpoint value used for QSAR modeling was expressed as pLC_{50} (mol/L) which refers to the negative logarithm of the concentration of chemical needed to kill the half of the population. For the larvicidal assay, the volume of stock solution is prepared as 20 ml of 1% and the stock solution is diluted in ethanol or other solvents. Three replicates are used for each concentration and the controls for the assay on different days. Each test solution contains 20 mosquito larvae and the mortality count is conducted after 24h which is a period of 12h light followed by 12 h dark. If the larvae don't reach the surface after probed with a needle, the larvae are count as dead. Concentration–mortality data are subjected to probit analysis (WHO, 2005).

Table 3.1. Experimental LC_{50} data compiled from the literature.

Endpoint	Chemicals	Larval age	Reference
LC_{50}	6 compounds (*Piper nigrum* Linn.)	Early 4th instar	Gulzar et al., 2013
	4 compounds (*Foeniculum vulgare*)	3rd instar	Rocha et al., 2015
	6 compounds (benzoquinone derivatives)		Sousa et al., 2010
	1 compound (tetradecanoic acid)		Sivakumar et al., 2011
	16 compounds (*Magnolia denudata*)		Wang et al., 2015
	55 compounds (terpenes, cyclic alcohols, etc.)		Santos et al., 2011 & Scotti et al., 2014

Average values taken for the chemicals assayed in different studies. For example, The LC_{50} data taken from Wang and colleagues (2015) and Santos and colleagues (2010; 2011) had common experimental data for some chemicals, so their average values were taken. For the study of Sousa and colleagues (2010), the purities of the chemicals were evaluated, and the chemicals have 90% purity, and more were included in the dataset and the others were included in external set for prediction. The final dataset includes 82 chemicals and the data compiled from the literature is shown in the Table 3.1. The range of pLC_{50} values is from 2.04 to 4.80 (mol/L).

3.2. Calculation of molecular descriptors:

The structures of molecules were drawn and geometrically optimized with SPARTAN v.16 (Wavefunction, 2016) using the semi-empirical PM6 method (Stewart, 2007). For the molecular descriptor calculations, the lowest energy conformations of molecular geometries were selected. Semi empirical molecular descriptors namely gaseous phase energy (*E*), highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), dipole moments, space filling (CPK) volume, space filling (CPK) area, polar surface area (PSA), accessible surface area, polar area, n-octanol-water partition coefficient (logP), hydrogen bond donor (HBD) count, hydrogen bond acceptor (HBA) count, polarizability, zero-point vibrational energy (ZPE) were calculated using SPARTAN 16.0 software (Wavefunction, 2016). Descriptors from DRAGON v.7, were also included in the descriptor pool. DRAGON software provides descriptors from 29 blocks including several topological and geometrical descriptors. As there were about 5000 descriptors in the pool, prior to modelling, to compute only those variables that are significant for the study, constant or near constant variables were excluded and uploaded into QSARINS 2.2.2 software (Gramatica et al. 2013, 2014; QSARINS 2017). After all these steps 935 descriptors left. The number of descriptors decreased further using "All Subset" and genetic algorithm (GA) tools of QSARINS. Finally, there was 57 descriptors left for modelling.

3.3. *In silico* modelling

The dataset was split into training (80%) and test (20%) to generate models by using QSARINS software. Three different splitting methods were used: (i) splitting by response, (ii) splitting by structure and (iii) random splitting. For different divisions, all subset procedure and genetic algorithm (GA)-based iterative facilities implemented in the QSARINS software were used for descriptor selection. Multiple Linear Regression (MLR) based on Ordinary Least Square (OLS) method implemented in QSARINS software was used for model development.

3.4. Applicability Domain

The Applicability Domain (AD) of generated QSAR models was shown with leverage approach (OECD, 2007) which reveals outliers in both the descriptors' and response spaces. The structural threshold was set at a critical hat value ($h^* = 3[p+1]/n$, where h^* is the critical hat value, p is the number of descriptors of the model, and n is the number of training compounds). The AD was visualized via Williams' plot which is the plot of the standardized residuals vs. hat values. Response

thresholds were set at ±3 standardized residuals. Compounds that are outside these ranges were considered as outliers.

3.5. Selection of the best model

In order to select the best model, multicriteria decision-making (MCDM) score (Keller et al. 1991) implemented in QSARINS 2.2.2 software was used. MCDM procedure is mainly about scoring the models between 0 to 1 (0 means worst and 1 means the best) regarding the internal and external validation parameters. The model with best MCDM score was selected for further evaluation. Of the generated best models, models with the highest MCDM score were also expected to fulfill the OECD validation requirements (OECD, 2007) besides statistical quality. Final models from each division were used to predict the larvicidal activity of an external set chemicals with no experimental mortality data.

3.6. Predictive Performance of QSAR models

Predictive performances of the generated models were tested via Insubria Graph which is utilized by QSARINS software. For this, larvicidal activities of an external set of chemicals with no experimental mortality data were predicted using the generated QSAR models. External dataset chemicals used in this study were retrieved from the literature and listed in the Appendix A. The chemicals are mostly plant extracts and found in the roots or leaves of different plants. Some of them are derivatives of the plant-based chemicals but used as repellants, so in the present study, the predictivity of these repellants was also evaluated. The external set was composed of 148 different chemicals and its composition is shown in Figure 3.1. The descriptors appearing in all QSAR models were calculated for the external set chemicals.

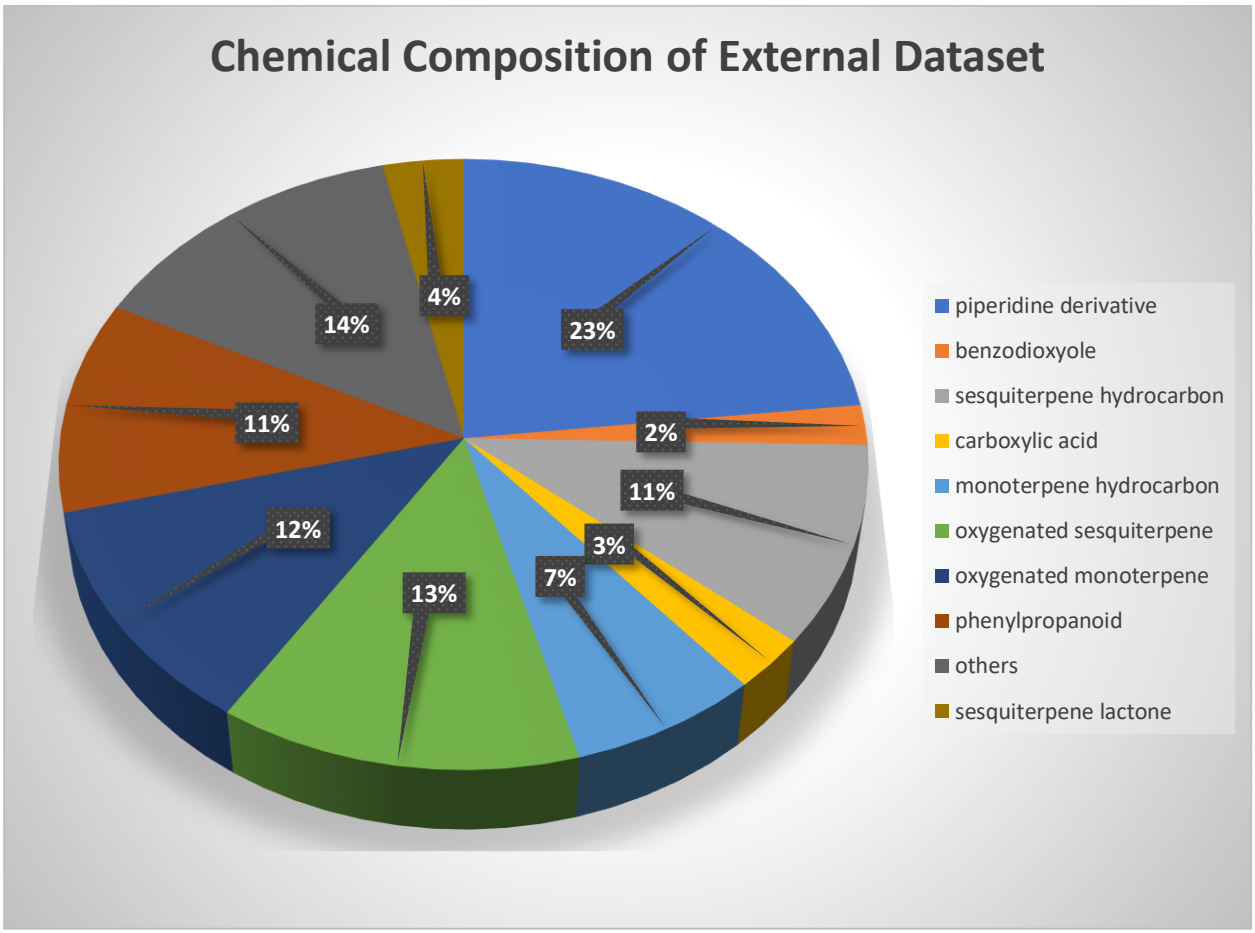

Figure 3.1. Chemical classes of 148 compounds in the external set.

3.7. Prediction of aquatic toxicity of larvicides

For algae toxicity prediction of larvicides, the comprehensive study conducted by Önlü and Saçan (2017a) which consists of quantitative structure–toxicity relationship (QSTR) models for the 72h algal toxicity data of hundreds of chemicals was used. For fish toxicity prediction of larvicides, the QSTR model developed by Önlü and Saçan (2017b) which assesses the cytotoxicity of different kind of chemicals on the rainbow trout (*Oncorhynchus mykiss*) liver cell line RTL-W1 and for the prediction of *Dugesia japonica* toxicity of larvicides, 5-descriptor QSTR model developed by Önlü and Saçan (2018) were used.

4. RESULTS AND DISCUSSION

4.1. Development of the QSAR model

4.1.1. Dataset

Chemicals in the dataset with their CAS number, physicochemical properties and experimental pLC_{50} values are listed in Table 4.1. Experimental pLC_{50} values of four chemicals were reported by different studies so their average values were taken and used in modeling.

Table 4.1. Physicochemical properties of chemicals used in the dataset.

LABEL	NAME	CAS NUMBER	log K_{ow}*	MOLECULAR WEIGHT**	pLC_{50} (Mol/L)	REFERENCE
1	carvacryl glycolic acid	NA	2.90	208.257	3.09	Scotti et al., 2013
2	1,8-cineole	470-82-6	1.86	154.253	2.04	Scotti et al., 2013
3	1,4-cineole	470-67-7	2.00	154.253	2.31	Scotti et al., 2013
4	carvacrol	499-75-2	3.37	150.221	3.47	Scotti et al., 2013
5	carvacryl benzoate	NA	5.24	254.329	3.66	Scotti et al., 2013
6	carvacryl acetate	6380-28-5	3.34	192.258	3.32	Scotti et al., 2013
7	carvacryl chloroacetate	NA	3.87	226.703	3.64	Scotti et al., 2013
8	2-hydroxy-3-methyl-6,-(1-methylethyl)-benzaldehyde	1665-99-2	3.11	178.231	3.43	Scotti et al., 2013
9	thymyl ethyl ether	NA	3.97	178.275	3.16	Scotti et al., 2013
10	thymoxyacetic acid	5333-40-4	2.9	208.257	2.65	Scotti et al., 2013
11	carvacryl propionate	NA	4.00	206.285	3.49	Scotti et al., 2013
12	carvacryl trichloroacetate	NA	5.23	295.593	3.59	Scotti et al., 2013
13	thymyl acetate	528-79-0	3.34	192.258	3.32	Scotti et al., 2013
14	thymyl chloroacetate	NA	3.87	226.703	3.66	Scotti et al., 2013
15	thymyl trichloroacetate	NA	5.23	295.593	3.85	Scotti et al., 2013
16	thymyl propionate	5451-69-4	4.00	206.285	3.49	Scotti et al., 2013
17	thymyl benzoate	NA	4.68	254.329	3.46	Scotti et al., 2013
18	2-hydroxy-6-methyl-3-(1-methylethyl)-benzaldehyde	1666-00-8	3.11	178.231	3.72	Scotti et al., 2013
19	5-norbornene-2-ol	13080-90-5	0.77	110.156	2.16	Scotti et al., 2013
20	5-norbornene-2,2-dimethanol	6707-12-6	0.61	154.209	2.29	Scotti et al., 2013
21	5-norbornene-2-endo-3-endodimethanol	699-97-8	0.33	154.209	2.04	Scotti et al., 2013
22	5-norbornene-2-exo-3-exo-dimethanol	699-95-6	0.33	154.209	2.33	Scotti et al., 2013
23	eugenyl acetate	93-28-7	2.55	206.241	3.28	Scotti et al., 2013
24	2-(2-methoxy-4-(2-propen-1-yl)) phenoxy acetic acid	NA	2.10	222.24	3.04	Scotti et al., 2013

Table 4.1. (Continued).

LABEL	NAME	CAS NUMBER	log K_{ow}*	MOLECULAR WEIGHT**	pLC_{50} (Mol/L)	REFERENCE
25	borneol	507-70-0	2.43	154.253	2.40	Scotti et al., 2013
26	catechol	120-80-9	1.25	110.112	2.66	Scotti et al., 2013
27	alpha-terpinene	99-86-5	2.96	136.238	3.76	Wang et al., 2014
28	terpineol	98-55-5	2.10	154.253	3.68	Wang et al., 2014
29	1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	155583-53-2	3.17	192.258	3.40	Scotti et al., 2013
30	eugenol	97-53-0	2.57	164.204	3.35	Scotti et al., 2013
31	phenol	108-95-2	1.64	94.113	2.69	Scotti et al., 2013
32	g-terpinene	99-85-4	2.96	136.238	3.54***	Santos et. al., 2011 & Wang et al., 2014
33	guaiacol	90-05-1	1.52	124.139	2.84	Scotti et al., 2013
34	1-benzoate-2-methoxy-4-(3-hydroxypropyl)-phenol	NA	3.51	286.327	3.28	Scotti et al., 2013
35	4-hydroxy-3-methoxy-benzenepropanol	2305-13-7	1.64	182.219	2.05	Scotti et al., 2013
36	isoborneol	124-76-5	2.43	154.253	2.41	Scotti et al., 2013
37	isopulegol	89-79-2	2.33	154.253	2.71	Scotti et al., 2013
38	thymol	89-83-8	3.37	150.221	2.59	Scotti et al., 2013
39	menthone	89-80-5	3.07	154.253	2.48	Santos et al., 2011
40	nonan-2-one	821-55-6	2.94	142.242	2.85	Scotti et al., 2013
41	undecan-2-one	112-12-9	3.78	170.296	3.51	Scotti et al., 2013
42	1,2-dimethoxy-4-(2-propen-1-yl)-benzene	93-15-2	2.83	178.231	3.24	Scotti et al., 2013
43	neo-isopulegol	122517-60-6	2.33	154.253	2.44	Santos et al., 2011
44	1,2-carvone oxide	36616-60-1	1.36	166.22	2.88	Santos et al., 2011
45	limonene oxide,cis	13837-75-7	1.83	152.237	2.47	Santos et al., 2011
47	p-cymene	99-87-6	2.65	134.222	3.51***	Santos et. al., 2011 & Wang et al., 2014
48	eugenyl propionate	7504-66-7	0.86	220.268	3.55	Scotti et al., 2013
49	R-carvone	6485-40-1	2.41	150.221	3.00	Santos et al., 2011
50	S-carvone	2244-16-8	2.41	150.221	3.08	Santos et al., 2011
51	R-limonene	5989-27-5	3.01	136.238	3.79***	Rocha et al.,2015 & Santos et al.,2011
52	S-limonene	138-86-3	3.01	136.238	3.83***	Rocha et al.,2015 & Santos et al.,2011
53	resorcinol	108-46-3	-0.62	110.112	2.28	Scotti et al., 2013
54	salicyl aldehyde	90-02-8	-0.56	122.123	2.95	Scotti et al., 2013
55	vanillin	121-33-5	-1.53	152.149	2.47	Scotti et al., 2013
56	2,6-dimethyl-p-benzoquinone	527-61-7	1.96	136.15	3.51	Sousa et al.,2010
55	vanillin	121-33-5	-1.53	152.149	2.47	Scotti et al., 2013
56	2,6-dimethyl-p-benzoquinone	527-61-7	1.96	136.15	3.51	Sousa et al.,2010
57	2,5-dimethyl-p-benzoquinone	137-18-8	1.96	136.15	3.38	Sousa et al.,2010
58	thymoquinone	490-91-5	2.71	164.204	3.53	Sousa et al.,2010

Table 4.1. (Continued).

LABEL	NAME	CAS NUMBER	log K_{ow}*	MOLECULAR WEIGHT**	pLC_{50} (Mol/L)	REFERENCE
59	pipilyasine	NA	5.51	279.468	3.99	Gulzar et al.,2013
60	pipzubedine	NA	718	335.576	4.18	Gulzar et al.,2013
61	pipyaqubine	NA	6.58	333.56	4.03	Gulzar et al.,2013
62	pellitorine	18836-52-7	4.26	237.387	4.07	Gulzar et al.,2013
63	pipericine	NA	7.41	339.608	4.13	Gulzar et al.,2013
64	piperine	94-62-2	2.95	285.343	4.45	Gulzar et al.,2013
65	(-)-camphene	5794-04-7	2.95	136.238	2.79	Santos et al., 2011
66	3-carene	13466-78-9	2.90	136.238	2.96	Santos et al., 2011
67	camphor	76-22-2	2.92	152.237	2.36	Scotti et al., 2013
68	menthol	89-78-1	2.75	156.269	2.59	Santos et al., 2011
69	tetradecanoic acid	544-63-8	4.94	228.376	3.96	Sivakumar et al., 2011
70	2,4-di-t-butylphenol	96-76-4	3.25	206.329	4.80	Wang et al., 2014
71	linoleic acid	60-33-3	5.97	280.452	4.59	Wang et al., 2014
72	nerolidol	7212-44-4	4.08	222.372	4.20	Wang et al., 2014
73	palmitic acid	57-10-3	5.77	256.43	3.88	Wang et al., 2014
74	methyl linolelaidate	2566-97-4	6.23	294.479	3.85	Wang et al., 2014
75	caryophyllene	87-44-5	4.48	204.357	3.53	Wang et al., 2014
76	geranic acid	4698--08-2	2.62	168.236	3.53	Wang et al., 2014
77	terpinen-4-ol	562-74-3	2.23	154.253	3.56	Wang et al., 2014
78	ethyl palmitate	628-97-7	6.37	284.484	3.73	Wang et al., 2014
79	humulene	6753-98-6	4.78	204.357	3.28	Wang et al., 2014
80	behenic acid	112-85-6	8.28	340.592	3.51	Wang et al., 2014
81	n-hexadecane	544-76-3	7.18	226.448	3.30	Wang et al., 2014
82	trans-anethole	4180-23-8	1.26	148.205	3.70	Rocha et al.,2015
83	estragole	140-67-0	1.31	148.205	3.50	Rocha et al.,2015

*n-octanol-water partition coefficients calculated by SPARTAN 10; ** MW from SPARTAN 10; *** Average values of experimental data.

4.1.2. Model Development

pLC_{50} dataset contains 82 chemicals (Table 4.1). The normality of the larvicidal activity data was evaluated using the Kolmogorov-Smirnov test using SPSS v.25 (IBM Corp, 2017) and the data showed a normal distribution ($p > 0.05$). The difference between minimum and maximum values in a data set should be at least 2 logarithmic unit in order to be used for QSAR modelling (Cronin et al., 2009). The range of logarithmic larvicidal activity is between (2.04) and (4.80). Thus, the pLC_{50} values were ordered and compounds with minimum and maximum activity values were left in the training set. Further splitting was made using the tool in QSARINS (v.2.2.2) software. The dataset was divided into training (80%) and test (20%) sets in order to build models. For each division, models with five descriptors were generated. The best models achieved in both response-based and structure-based divisions rather than random divisions by using QSARINS software. Therefore, these

models were selected for further analysis. The label numbers of test set chemicals for the selected divisions were given in Table 4.2.

Numerous divisions and models were generated. After passing the MCDM criteria, 5-descriptor models were selected regarding their high R^2 and Q^2_{LOO} values, the minimum number of structural outliers in Williams plot and higher prediction performance for the external set chemicals. The selected models were listed together with their fit and internal validation parameters, and external validation parameters in Table 4.3. and 4.4, respectively.

R^2 and Q^2_{LOO} values were very similar to each other; R^2 range is between 0.707 and 0.778, and Q^2_{LOO} range is between 0.658 and 0.733. The developed models were also confirmed with their performances on the test sets by applying and comparing different external validation criteria. $R^2_{test,}$ (the external determination coefficient), Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, r^2_m metrics, *CCC* (the concordance correlation coefficient), CCC_{test}. The *RMSE* (root mean squared of errors) were used to measure and compare prediction accuracy in the training ($RMSE_{Tr}$) and in the test sets ($RMSE_{test}$). R^2_{test} is a commonly used regression-based metric regarding the experimental data points for the test set chemicals which were not used during modelling step versus those calculated by the model equation, the more the predicted values match the experimental ones, the better the model performance (Gramatica et al., 2011). In this study, the R^2_{test} values range from 0.765 to 0.832 shows the good predictivity performance of the models.

All models satisfied the Golbraikh and Tropsha (2003) criteria. *CCC* is an additional criterion for the external validation suggested by Lin (1989) which verifies the agreement of experimental and predicted data and the suggested cutoff value for the *CCC* was 0.80. However, the following threshold values by Chirico and Gramatica (2012) were attained for the relevant parameters. For the developed models, the average $\bar{r}^2_m$ and Δr_m^2 were also in good agreement with the suggested limits, for r^2_m 0.661 to 0.753 (>0.5) and for Δr_m^2 0.001 to 0.192 (<0.2).

I. CCC_{test}= 0.85
II. Q^2_{Fn} = 0.70
III. $\bar{r}^2_m$ = 0.65
IV. Δr_m^2 = 0.20

The predictive performances of all models were further evaluated with *MAE*-based criteria (Roy et al., 2016). Only four of the generated models were found to be in “good” category based on the

MAE criteria Table 4.4. It should be noted that although a model passes all of the internal and external validation criteria, it may fail to be valid in terms of *MAE*-based criteria. The predicted vs observed graphs for these four models were given in Appendix B.

Table 4.2. The label numbers of test set chemicals for three divisions used in the QSAR modeling.

Division no	Test Set Compounds* (n_{test}/n_{Tr}= 1/4)
1	3,9,14,16,21,27,29,43,48,63,65,66,68,73,80,81
2	3,11,12,24,29,48,51,55,58,61,62,63,65,66,75,79
3	1,5,11,20,32,36,37,38,54,56,57,62,71,74,78,79

*Compound numbers refer to label numbers given in Table 4.1

Fit, internal and external validation parameters of the models are given in Table 4.3 and 4.4, respectively.

Table 4.3. Fit and internal validation parameters of the generated models.

Label	Descriptors	R^2	R^2_{adj}	$RMSE_{Tr}$	CCC_{Tr}	F	Q^2_{LOO}	$RMSE_{CV}$	CCC_{CV}
	Division 1								
D1M1	Chi_D/Dt SpPosA_B(v) GATS7e SpMin6_Bh(s) Mor30s	0.762	0.742	0.309	0.865	38.427	0.711	0.340	0.837
D1M2	GATS7e Mor25m HATS1s Ui SsssCH	0.761	0.741	0.309	0.864	38.228	0.711	0.341	0.837
D1M3	WiA_D/Dt MATS7e P_VSA_i_2 Mor25m	0.707	0.688	0.343	0.828	36.834	0.658	0.370	0.801
D1M4	Chi_D/Dt SpPosA_B(p) GATS7e SpMin6_Bh(s) Mor30v	0.753	0.733	0.314	0.859	36.688	0.697	0.349	0.828
D1M5	ATSC8p GATS7e Mor30s HATS1s TDB03v	0.746	0.725	0.319	0.855	35.333	0.699	0.348	0.828
	Division 2								
D2M1	GATS7e Mor25m HATS1s Ui nCrt	0.779	0.761	0.303	0.8761	42.428	0.733	0.333	0.851
D2M2	**GATS7e HATS1s H-046 Ui Mor30s**	**0.760**	**0.740**	**0.315**	**0.8641**	**38.134**	**0.708**	**0.348**	**0.834**
	Division 3								
D3M1	GATS7e HATS1s H-046 Ui Mor30s	0.772	0.753	0.300	0.8718	40.817	0.722	0.331	0.843
D3M2	GATS7e Mor25m HATS1s Ui nCrt	0.778	0.760	0.296	0.8757	42.275	0.733	0.325	0.850
D3M3	IC3 GATS7e P_VSA_i_2 Mor25m nCt	0.757	0.737	0.310	0.8621	37.507	0.695	0.347	0.827

Table 4.4. External validation parameters of the generated QSAR models.

Label	Descriptors	R^2_{test}	$RMSE_{test}$	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	CCC_{test}	$r^2_{m\ av.}$	Δr^2_m	k'	k	$(R^2 - R_0'^2)/R^2$	$(R^2 - R_0^2)/R$
Division 1													
D1M1	Chi_D/Dt SpPosA_B(v) GATS7e SpMin6_Bh(s) Mor30s	0.829	0.242	0.832	0.828	0.853	0.907	0.753	0.130	0.992	1.002	0.034	0.0002
D1M2	GATS7e Mor25m HATS1s Ui SsssCH	0.792	0.283	0.772	0.767	0.800	0.885	0.706	0.040	0.978	1.014	0.008	0.0227
D1M3	WiA_D/Dt MATS7e P_VSA_i_2 Mor25m	0.793	0.277	0.782	0.777	0.808	0.873	0.661	0.192	1.014	0.979	0.104	0.0025
D1M4	Chi_D/Dt SpPosA_B(p) GATS7e SpMin6_Bh(s) Mor30v	0.791	0.274	0.786	0.781	0.812	0.877	0.675	0.188	1.010	0.982	0.089	0.0009
D1M5	ATSC8p GATS7e Mor30s HATS1s TDB03v	0.790	0.278	0.779	0.774	0.806	0.880	0.700	0.145	1.015	0.978	0.053	0.0006
Division 2													
D2M1	GATS7e Mor25m HATS1s Ui nCrt	0.765	0.263	0.764	0.746	0.833	0.874	0.670	0.002	1.003	0.990	0.020	0.0195
D2M2	**GATS7e HATS1s H-046 Ui Mor30s**	**0.826**	**0.228**	**0.823**	**0.809**	**0.875**	**0.907**	**0.752**	**0.038**	**1.005**	**0.990**	**0.005**	**0.0155**
Division 3													
D3M1	GATS7e HATS1s H-046 Ui Mor30s	0.773	0.305	0.748	0.744	0.764	0.875	0.682	0.001	1.014	0.978	0.017	0.0184
D3M2	GATS7e Mor25m HATS1s Ui nCrt	0.794	0.294	0.767	0.763	0.782	0.883	0.709	0.026	1.020	0.973	0.019	0.0103
D3M3	IC3 GATS7e P_VSA_i_2 Mor25m nCt	0.788	0.285	0.781	0.777	0.795	0.883	0.699	0.101	1.009	0.983	0.039	0.003

4.1.3. Comparison of Applicability Domain of the Models

In order to select the best QSAR model, the applicability domain of each model was tested to predict pLC_{50} of 148 chemicals from various classes with no experimental data for 24h lethal concentration for the 3rd and the early 4th instar *Aedes aegypti* larvae. The number of chemicals that each model could predict out of 148 were given in Table 4.5 with corresponding structural coverage.

Table 4.5. Predictive performance and *MAE*-based criteria of the generated models.

Division	Model label	Number of chemicals in AD (out of 148)	Structural coverage (%)	*MAE*-based criteria
1	D1M1	118	79.7	Good
	D1M2	124	83.7	Moderate
	D1M3	139	93.9	Good
	D1M4	114	77.0	Moderate
	D1M5	131	88.5	Moderate
2	D2M1	127	85.8	Good
	D2M2	**140**	**94.6**	**Good**
3	D3M1	139	93.9	Bad
	D3M2	129	87.1	Moderate
	D3M3	128	86.4	Moderate

By comparing all the statistical parameters given in Table 4.3 and Table 4.4 and the structural coverage ratios of the generated models, the best model is decided and written in bold (D2M2). The selected model is given in Eq. 4.1, together with the 95% confidence interval for the coefficients within parenthesis:

$$pLC_{50} = 2.7384\ (\pm 0.5378) + 0.2933\ (\pm 0.1345)\ GATS7e - 1.2660\ (\pm 0.7213)\ HATS1s + 0.0248\ (\pm 0.0155)\ H\text{-}046 + 0.4556\ (\pm 0.1404)\ Ui - 0.1680\ (\pm 0.1344)\ Mor30s \quad (4.1)$$

This 5-descriptor model has high R^2 (R^2= 0.760, R^2_{adj} = 0.740) and low $RMSE_{Tr}$= 0.315 showing that the chosen model is satisfactory. Also, the robustness and the stability of the model is proven by Q^2_{LOO} (0.708). The possibility of chance correlation is eliminated with very low R^2_{Yscr} (0.078) and Q^2_{Yscr} (-0.119) values. Predictive performance of the model is shown by high R^2_{test} and low $RMSE_{test}$ values (0.826 and 0.228, respectively). External validation parameters Q^2_{F1}, Q^2_{F2}, Q^2_{F3} and CCC_{test} for the selected model have higher values than the corresponding literature threshold values which prove the predictive performance of the model. Other external validation parameters shown in Table

4.4 prove that the model is satisfactory. Moreover, the model is "acceptable" regarding with the criteria presented by Golbraikh and Tropsha (2002). The predictive performance of the model is classified as "good" with values of *MAE* (95% of the data) = 0.134, $3\sigma = 0.469$ and training set range as 2.760.

Chemicals used for the QSAR modeling with their training/test set status in modelling, experimental and predicted pLC_{50} values, hat and descriptor values are given in Appendix C.

The 5-descriptor MLR model labelled as D2M2 and highlighted in Tables 4.3. and 4.4. has no response and structural outliers. Additionally, its structural coverage is 94.6 % for the external set chemicals (Table 4.5).

Figure 4.1. shows the experimental pLC_{50} values from versus predicted pLC_{50} values from Eq. 4.1. As it can be seen in Figure 4.1, data points are homogenously distributed to the optimal line indicating that the model has a good fit:

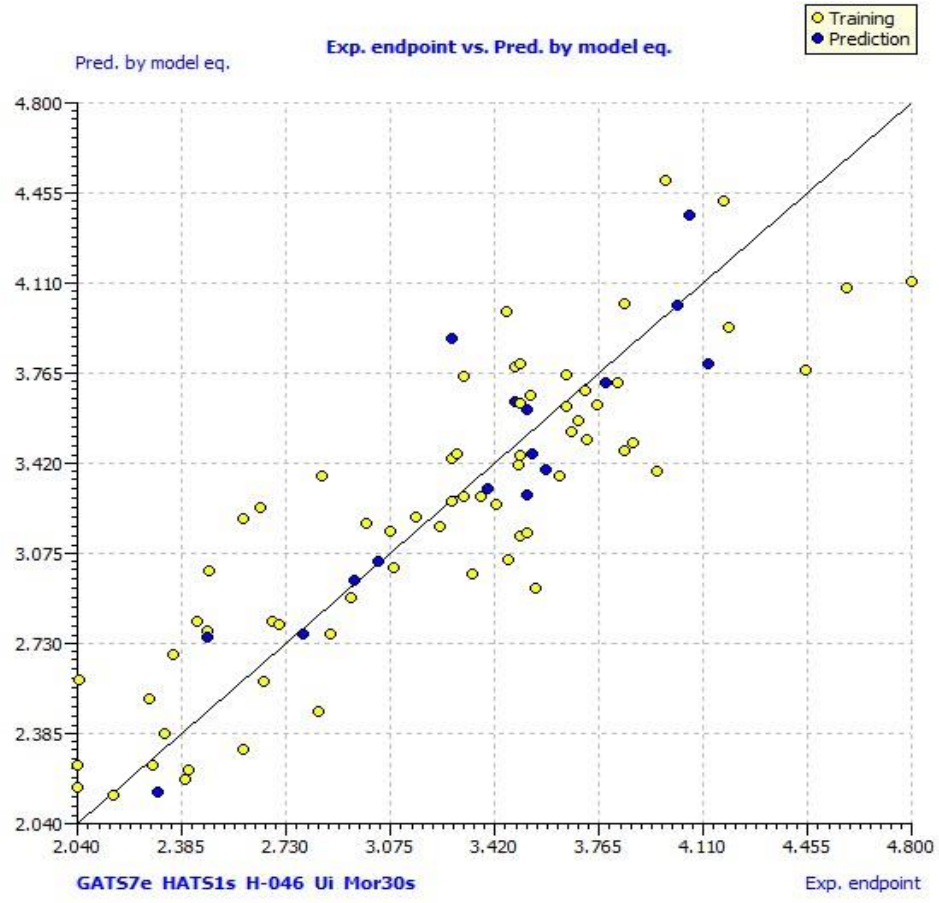

Figure 4.1. Plot of predicted pLC_{50} values from Eq. 4.1 versus experimental pLC_{50} values.

The applicability domain of model was visualized via Williams plot which is the plot of standard residuals versus hat values. Response thresholds were set at ±3 standardized residuals. Figure 4.2 shows the applicability domain of the model and there are no response and structure outliers. None of the compounds has higher hat value than the critical hat value (h^*= 0.273) of the model. So, all of the compounds belong to the applicability domain and the predicted values are expected to be reliable.

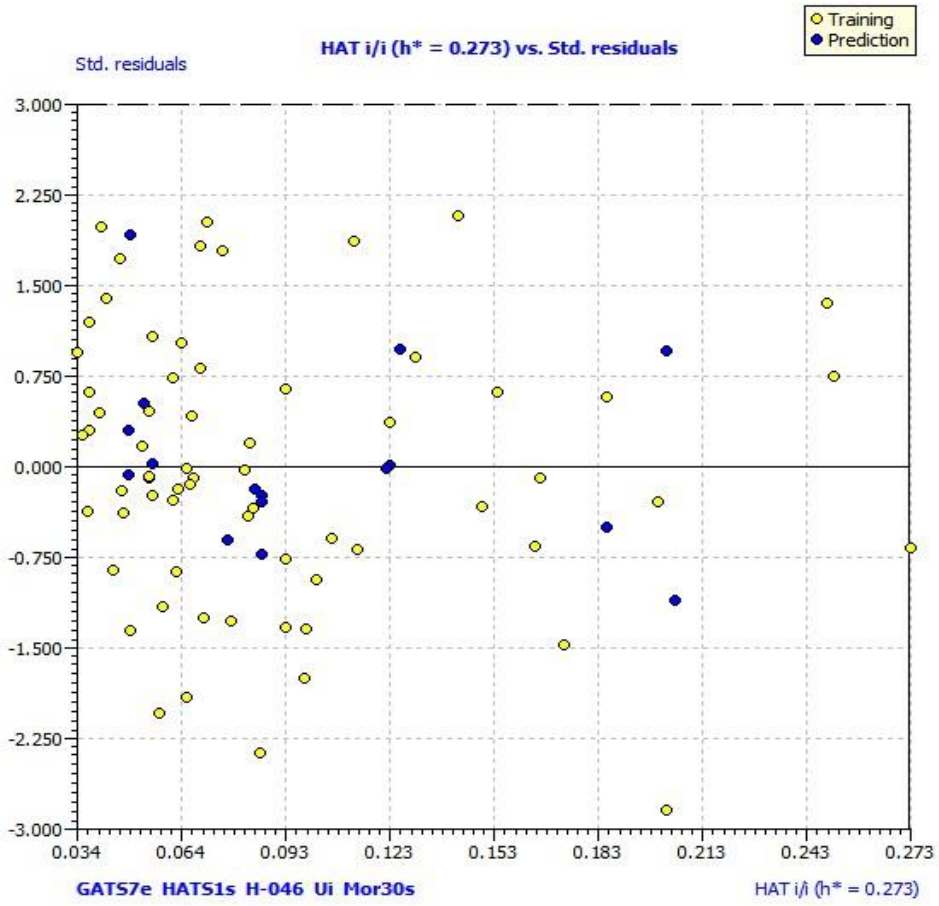

Figure 4.2. Williams plot of the model.

Insubria graph was to assess the reliability of the predictions of compounds lacking experimental response (Figure 4.3). In the external set, there are 148 chemicals. Predicted pLC_{50} values and calculated descriptors of external set compounds, are provided in Appendix D (D1). The model had 94.6% of structural coverage as mentioned above in Table 4.5. The name of the compounds regarding with labels are also provided in Appendix D.

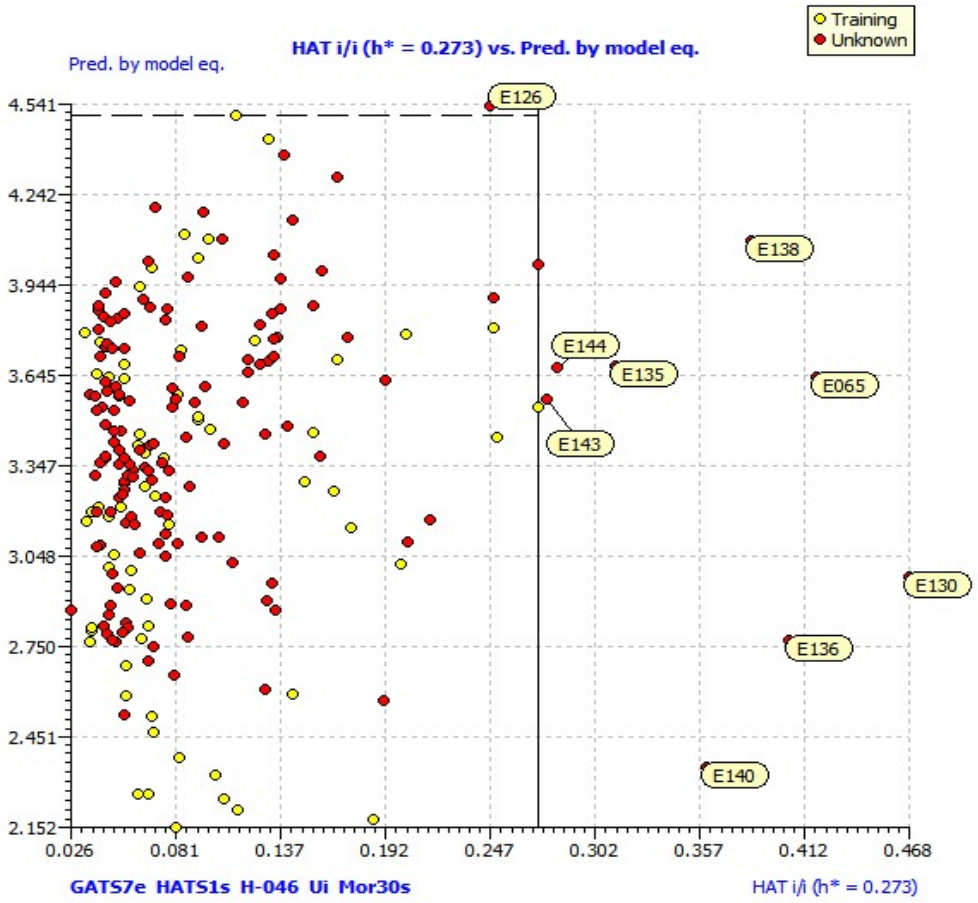

Figure 4.3. Insubria graph of model (Eq. 4.1). Predicted pLC_{50} values for training and external (148 chemicals) set chemicals from Eq. 4.1 versus their hat values.

The most and the least ten active compounds predicted by Eq. 4.1 is listed in Table 4.6. The most active compound was found as "1-Undec-10-enoyl-4-benzyl-piperidine" with the highest pLC_{50} (4.37) value. The predicted value of "E126: guineensine" cannot be reliable because the its predicted activity value is slightly higher than the response range of model (Eq. 4.1).

Table 4.6. The more and the less active chemicals from the external set predicted by Eq. 4.1.

	Name	Chemical Class	Predicted pLC_{50} value from Eq. 4.1 (mol/L)
More active	1-undec-10-enoyl-4-benzyl-piperidine	Piperidine derivative	4.37
	pipercide	Alkaloid	4.30
	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	Piperidine derivative	4.20
	3-benzyl-1-undec-10-enoyl-piperidine	Piperidine derivative	4.18
	retrofractamide A	Alkaloid	4.15
	2-benzyl-1-undec-10-enoyl-piperidine	Piperidine derivative	4.09
	alpha-bisabolol	Oxygenated sesquiterpene	4.04
	1-octanoyl-3-benzyl-piperidine	Piperidine derivative	4.02
	alpha-eudesmol	Oxygenated sesquiterpene	4.01
	tectoquinone	Anthraquinone	3.99
Less active	citronellic acid	Carboxylic acid	2.52
	p-menthane-3,8-diol	Oxygenated monoterpene	2.57
	patchouli alcohol	Oxygenated sesquiterpene	2.61
	verbenone	Oxygenated monoterpene	2.65
	Z, E- nepetalactone	Oxygenated monoterpene	2.70
	ascaridole	Oxygenated monoterpene	2.75
	piperitone oxide	Oxepane	2.76
	(+)-camphene	Monoterpene hydrocarbon	2.77
	3,4,5-trimethoxy toluene	Benzene derivative	2.78
	fenchene	Oxygenated monoterpene	2.79

The model name “D3M1” is also worth to consider because of high structural coverage, but it can be seen that it has the same descriptors with the best model, also the coefficients of the descriptors don’t change much in the equations 4.1 and 4.2. The equation for model “D3M1” is:

$$pLC_{50} = 2.8076\ (\pm 0.4726) + 0.2809\ (\pm 0.1271)\ GATS7e - 1.3987\ (\pm 0.6507)\ HATS1s + 0.0238\ (\pm 0.0133)\ H\text{-}046 + 0.4723\ (\pm 0.1384)\ Ui - 0.1636\ (\pm 0.1405)\ Mor30s \quad (4.2)$$

So, in order to include all the information given by the main dataset, a full model which has the same descriptors with the chosen model is generated and this full model is re-calculated with the combination of training and test set together. At the end the equation of full model and the statistical parameters are shown below (Eq. 4.3):

$$pLC_{50} = 2.7291\ (\pm 0.4358) + 0.2713\ (\pm 0.1111)\ GATS7e - 1.2273\ (\pm 0.5846)\ HATS1s + 0.0257\ (\pm 0.0124)\ H\text{-}046 + 0.4493\ (\pm 0.1195)\ Ui - 0.1654\ (\pm 0.1143)\ Mor30s \quad (4.3)$$

n=231 R^2= 0.7703 Q^2_{LOO}= 0.7325 Q^2_{LMO}= 0.7225 R^2_{Yscr}= 0.0603 $RMSE_{Tr}$= 0.3001 $RMSE_{CV}$= 0.3238 MAE_{Tr}= 0.2358 MAE_{CV}= 0.2548 TSR=2.760

Chemicals used for the QSAR modeling with their experimental and predicted pLC_{50} values by Eq. 4.3, hat and descriptor values are given in Table 4.7.

The descriptor ranges for full model are as follows: GATS7e: 0 to 2.854, HATS1s: 0.146 to 1.549, H-046: 0 to 41, Ui: 0 to 3 and Mor30s: -1.269 to 1.964, model prediction range pLC_{50} is 2.17 to 4.47.

Table 4.7. Chemicals used for the QSAR model, their experimental and predicted pLC_{50} values from Eq. 4.3, hat values and descriptor values.

Name	**Exp. pLC_{50} (mol/L)**	**Pred. pLC_{50} by Eq. 4.3**	**Hat values (*h**=0.219)**	**GATS7e**	**HATS1s**	**H-046**	**Ui**	**Mor30s**
carvacryl glycolic acid	3.09	3.01	0.151	1.393	0.883	10	2.322	1.860
1,8-cineole	2.04	2.19	0.144	0.000	0.389	4	0.000	0.977
1,4-cineole	2.31	2.17	0.143	0.000	0.374	6	0.000	1.490
carvacrol	3.47	3.06	0.040	0.160	0.680	10	2.000	0.174
carvacryl benzoate	3.66	3.75	0.097	0.675	0.503	10	3.000	0.896
carvacryl acetate	3.32	3.29	0.042	0.696	0.595	10	2.322	1.197
carvacryl chloroacetate	3.64	3.36	0.033	1.060	0.621	10	2.322	1.144
2-hydroxy-3-methyl-6,-(1-methylethyl)-benzaldehyde	3.43	3.27	0.135	0.097	0.769	10	2.322	-0.985
thymyl ethyl ether	3.16	3.21	0.040	0.181	0.498	10	2.000	0.627
thymoxyacetic acid	2.65	3.24	0.054	1.006	0.793	10	2.322	0.499
carvacryl propionate	3.49	3.65	0.041	0.638	0.509	13	2.322	-0.003
carvacryl trichloroacetate	3.59	3.38	0.056	1.537	0.765	10	2.322	0.746
thymyl acetate	3.32	3.72	0.033	1.648	0.570	10	2.322	0.299
thymyl chloroacetate	3.66	3.61	0.041	1.696	0.666	10	2.322	0.351
thymyl trichloroacetate	3.85	3.44	0.073	1.830	0.772	10	2.322	0.783
thymyl propionate	3.49	3.76	0.025	1.479	0.521	13	2.322	0.614
thymyl benzoate	3.46	3.97	0.054	1.778	0.613	9	3.000	0.412
2-hydroxy-6-methyl-3-(1-methylethyl)-benzaldehyde	3.72	3.67	0.133	1.722	0.771	10	2.322	-0.750
5-norbornene-2-ol	2.16	2.17	0.066	0.000	0.721	3	1.000	1.209
5-norbornene-2,2-dimethanol	2.29	2.28	0.056	0.000	0.700	6	1.000	1.135
5-norbornene-2-endo-3-endodimethanol	2.04	2.28	0.052	0.000	0.715	4	1.000	0.735
5-norbornene-2-exo-3-exo-dimethanol	2.33	2.40	0.073	0.000	0.732	4	1.000	-0.126
eugenyl acetate	3.28	3.26	0.050	0.984	0.659	2	2.585	0.835
2-(2-methoxy-4-(2-propen-1-yl))-phenoxy acetic acid	3.04	3.03	0.094	1.123	0.777	2	2.585	1.582
borneol	2.40	2.23	0.095	0.000	0.467	14	0.000	1.676
catechol	2.66	2.59	0.046	0.000	0.767	0	2.000	0.532

Table 4.7. (Continued).

Name	**Exp. pLC_{50} (mol/L)**	**Pred. pLC_{50} by Eq. 4.3**	**Hat values (*h**=0.219)**	**GATS7e**	**HATS1s**	**H-046**	**Ui**	**Mor30s**
alpha-terpinene	3.76	3.62	0.038	1.625	0.446	14	1.585	0.443
terpineol	3.68	3.49	0.215	2.854	0.457	9	1.000	0.782
1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	3.40	3.31	0.067	0.687	0.514	2	2.322	0.416
eugenol	3.35	2.99	0.043	1.041	0.817	2	2.322	0.684
phenol	2.69	2.82	0.054	0.000	0.632	0	2.000	0.180
g-terpinene	3.54	3.65	0.045	1.625	0.441	14	1.585	0.250
guaiacol	2.84	2.48	0.057	0.000	0.834	0	2.000	0.737
1-benzoate-2-methoxy-4-(3-hydroxypropyl)-phenol	3.28	3.41	0.196	1.131	0.563	2	3.000	1.964
4-hydroxy-3-methoxy-benzenepropanol	2.05	2.59	0.106	1.030	0.941	2	2.000	1.272
isoborneol	2.41	2.27	0.089	0.000	0.460	14	0.000	1.524
isopulegol	2.71	2.81	0.032	0.217	0.517	11	1.000	0.422
thymol	2.59	3.21	0.033	0.224	0.581	10	2.000	0.092
menthone	2.48	3.02	0.035	0.180	0.392	15	1.000	0.644
nonan-2-one	2.85	3.36	0.061	1.217	0.348	13	1.000	0.329
undecan-2-one	3.51	3.44	0.049	0.986	0.290	17	1.000	0.487
1,2-dimethoxy-4-(2-propen-1-yl)-benzene	3.24	3.16	0.036	0.795	0.636	2	2.322	0.547
neo-isopulegol	2.44	2.82	0.031	0.217	0.502	11	1.000	0.462
1,2-carvone oxide	2.88	2.77	0.032	0.162	0.649	4	1.585	0.102
limonene oxide,cis	2.47	2.78	0.052	0.298	0.496	6	1.000	0.133
p-cymene	3.51	3.62	0.032	1.577	0.512	10	2.000	0.355
eugenyl propionate	3.55	3.43	0.066	1.180	0.573	5	2.585	1.197
R-carvone	3.00	3.19	0.031	0.280	0.593	9	2.000	0.052
S-carvone	3.08	3.16	0.029	0.280	0.610	9	2.000	0.105
R-limonene	3.79	3.70	0.069	2.031	0.466	13	1.585	0.321
S-limonene	3.83	3.69	0.068	2.031	0.466	13	1.585	0.338
resorcinol	2.28	2.53	0.058	0.000	0.874	0	2.000	0.118
salicyl aldehyde	2.95	2.91	0.057	0.000	0.741	0	2.322	-0.292
vanillin	2.47	2.74	0.147	1.454	1.052	0	2.322	0.796

Table 4.7. (Continued).

Name	**Exp. pLC_{50} (mol/L)**	**Pred. pLC_{50} by Eq. 4.3**	**Hat values (*h**=0.219)**	**GATS7e**	**HATS1s**	**H-046**	**Ui**	**Mor30s**
2,6-dimethyl-p-benzoquinone	3.51	3.15	0.146	0.000	0.799	6	2.322	-1.269
2,5-dimethyl-p-benzoquinone	3.38	3.30	0.127	0.000	0.677	6	2.322	-1.245
thymoquinone	3.53	3.31	0.071	0.122	0.668	10	2.322	-0.416
pipilyasine	3.99	4.47	0.090	2.156	0.252	24	2.000	0.291
pipzubedine	4.18	4.40	0.097	1.873	0.248	32	2.000	1.499
pipyaqubine	4.03	4.02	0.086	0.816	0.242	27	2.000	1.340
pellitorine	4.07	4.32	0.102	2.577	0.299	18	2.000	0.589
pipericine	4.13	3.82	0.150	0.741	0.233	39	1.000	1.640
piperine	4.45	3.75	0.158	0.727	0.388	2	2.807	0.037
(-)-camphene	2.79	2.78	0.037	0.000	0.461	14	1.000	1.113
3-carene	2.96	2.98	0.042	0.000	0.407	15	1.000	0.463
camphor	2.36	2.70	0.044	0.000	0.493	14	1.000	1.353
menthol	2.59	2.35	0.087	0.191	0.498	15	0.000	1.240
tetradecanoic acid	3.96	3.39	0.049	0.847	0.315	25	1.000	1.613
2,4-di-*t*-butylphenol	4.80	4.07	0.068	2.338	0.438	18	2.000	0.657
linoleic acid	4.59	4.09	0.069	0.882	0.254	25	2.000	0.646
nerolidol	4.20	3.91	0.048	1.814	0.362	15	2.000	0.911
palmitic acid	3.88	3.51	0.070	0.796	0.275	29	1.000	1.753
methyl linolelaidate	3.85	4.03	0.065	0.855	0.273	25	2.000	0.832
caryophyllene	3.53	3.62	0.035	1.128	0.350	21	1.585	1.432
geranic acid	3.53	3.15	0.059	1.021	0.753	13	2.000	0.954
terpinen-4-ol	3.56	2.95	0.047	0.210	0.452	11	1.000	0.075
ethyl palmitate	3.73	3.52	0.069	0.763	0.264	29	1.000	1.719
humulene	3.28	3.89	0.035	1.146	0.347	20	2.000	0.820
behenic acid	3.51	3.82	0.184	0.725	0.190	41	1.000	2.246
n-hexadecane	3.30	3.47	0.122	0.955	0.146	34	0.000	1.286
trans-anethole	3.70	3.55	0.064	1.355	0.502	3	2.322	0.292
estragole	3.50	3.38	0.046	1.549	0.656	2	2.322	0.303

Full model (Eq. 4.3) then applied to the external data seta and the applicability of full model has higher structural coverage 95.3% compared to the selected model. Predicted pLC_{50} values and calculated descriptors of external set compounds, are provided in Appendix D (Table D2). The Insubria graph of full model shown in Figure 4.4:

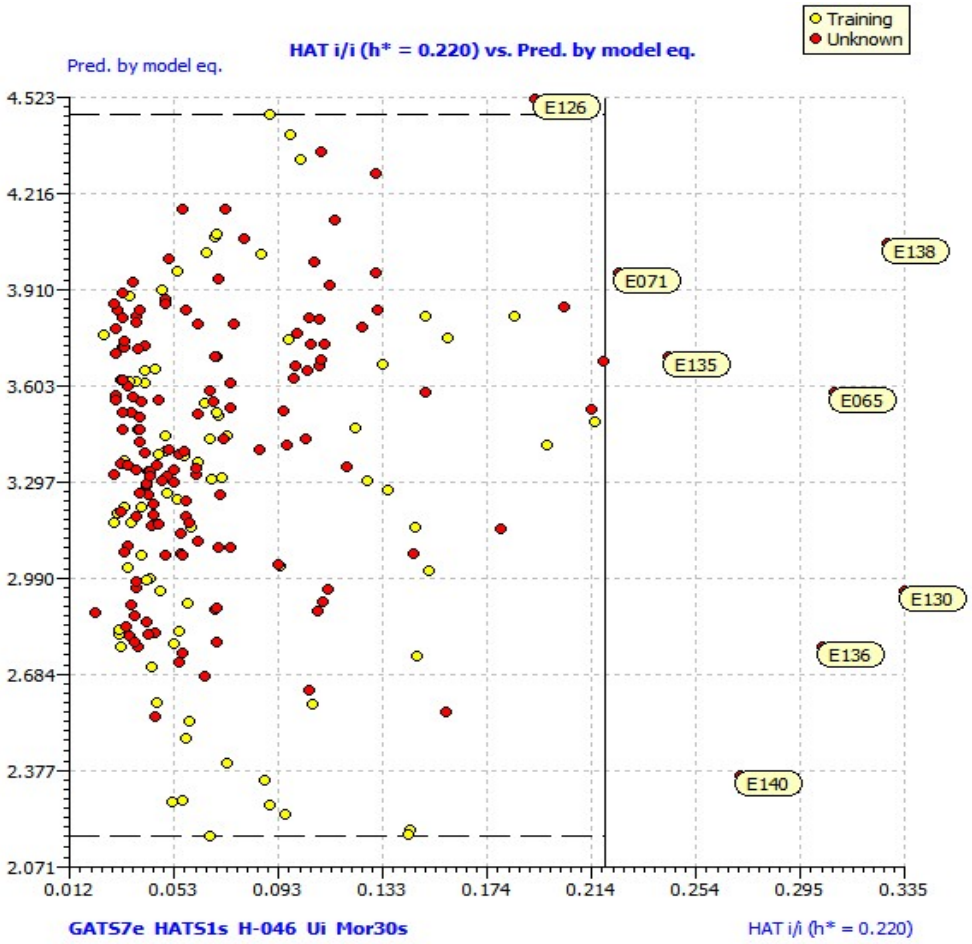

Figure 4.4. Insubria graph of full model (Eq. 4.3).

Full model has made predictions on the three chemicals whose purities lower than 90%. The prediction results show lower values for the larvicidal activity of these compounds. The experimental pLC_{50} values for para-benzoquinone, 2-methyl-para-benzoquinone and 2-isopropyl-para-benzoquinone are 3.07, 3.30 and 3.65, respectively. Whereas, the predicted pLC_{50} values are 2.89, 2.91 and 2.88, respectively.

The most and the least toxic compounds predicted by full model (Eq. 4.3) is shown in Table 4.8. Again “E126: guineensine” is within the structural AD of full model, but its prediction may not be reliable.

Table 4.8. The most and the least larvicidal activity of 10 chemicals from the external set predicted by full model (Eq. 4.3).

	Name	Chemical Class	Predicted pLC_{50} value from Eq. 4.3 (mol/L)
More active	1-undec-10-enoyl-4-benzyl-piperidine	Piperidine derivative	4.34
	pipercide	Alkaloid	4.28
	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	Piperidine derivative	4.17
	3-benzyl-1-undec-10-enoyl-piperidine	Piperidine derivative	4.17
	retrofractamide A	Alkaloid	4.13
	2-benzyl-1-undec-10-enoyl-piperidine	Piperidine derivative	4.07
	1-octanoyl-3-benzyl piperidine	Piperidine derivative	4.01
	alpha-bisabolol	Oxygenated sesquiterpene	4.00
	tectoquinone	Anthraquinone	3.96
	alpha-eudesmol	Oxygenated sesquiterpene	3.96
Less active	citronellic acid	Carboxylic acid	2.55
	p-menthane-3,8-diol	Oxygenated monoterpene	2.56
	patchouli alcohol	Oxygenated sesquiterpene	2.63
	verbenone	Oxygenated monoterpene	2.67
	Z,E- nepetalactone	Oxygenated monoterpene	2.72
	ascaridole	Oxygenated monoterpene	2.75
	piperitone oxide	Oxepane	2.77
	(+)-camphene	Monoterpene hydrocarbon	2.78
	3,4,5- trimethoxy toluene	Benzene derivative	2.79
	fenchene	Oxygenated monoterpene	2.81

Both alkaloids and piperidine derivatives were found to be effective against 3rd and early 4th instar *Aedes aegypti* larvae regarding full model. Larvicidal activity of pipercide was examined by Park and colleagues (2002) and 48h- pLC_{50} value of this compound was found as 6.55 (mol/L). Piperidine derivatives were examined as adulticide to female *Aedes aegypti* by Pridgeon and colleagues. (2007) but their larvicidal activities haven't been examined yet. This study may provide further emphasis on *Piper nigrum* because these compounds that are found to be effective against *Aedes aegypti* larvae can be extracted via this fruit. The descriptors appearing in full model, corresponding types and their coefficients are shown in Table 4.9.

The order of descriptors' importance is Ui > HATS1s > H-046 > GATS7e > Mor30s.

Table 4.9. Descriptors appeared in full model (Eq. 4.3)

Descriptor	**Type**	**Meaning of descriptor**	**Standardized coefficient**
GATS7e	2D autocorrelations	Geary autocorrelation of lag 7 weighted by Sanderson electronegativity	0.315
HATS1s	GETAWAY descriptors	leverage-weighted autocorrelation of lag 1 / weighted by I-state	-0.379
H-046	Atom-centered fragments	H attached to $C^0(sp^3)$ no X attached to next C	0.372
Ui	Molecular properties	unsaturation index	0.538
Mor30s	3D-MoRSE descriptors	signal 30 / weighted by I-state	-0.180

Unsaturation index (Ui) is the most important descriptor in the model regarding the highest standardized coefficient it has in the model. The positive regression coefficient of this descriptor indicates that with an abundance of unsaturated bonds in the structure the pLC_{50} value increases. Ui can be described with the equation below (Eq. 4.4):

$$UI = \log_2 (1+ b) \quad (4.4)$$

where b is calculated with the equation (Eq. 4.5):

$$b = (2N_C + 2N_H - N_x + N_N + N_P + 2(N_{O\text{-}S} - N_{SO_3})) / 2 - C \quad (4.5)$$

N_C, N_H, N_X, N_N, N_P, and C are the number of carbon atoms, hydrogen, halogen, nitrogen, phosphorous, and independent cycles, respectively. N_{O-S} and N_{SO_3} are the number of oxygen atoms bonded to sulfur and the number of SO_3 groups, respectively. When there is no sulfur atom in the compound, this index can be calculated from the chemical formula; otherwise the index calculated replacing b with b* (Eq. 4.6) which is the more general form of UI:

$$b^* = \sum_b (\pi^*_{ij})_b - B \tag{4.6}$$

where π^* is the conventional bond order (the conventional bond order π^* is defined as being equal to 1, 2, 3, and 1.5; for single, double, triple, and aromatic bonds, respectively). For saturated compounds, b* = UN = 0 (Todeschini and Consonni, 2008). Unsaturation was highlighted by Lomonaco and colleagues (2008) as to explain the larvicidal activity. According to the study, double bond increases the lipophilic character which makes easier to go through larvae cuticle, thus larvicidal activity increases with unsaturation. This descriptor was used in another study done by De and Roy (2018) to investigate the persistency, bioaccumulation and toxicity (PBT) index of dioxins, PAHs, PCBs, pesticides and various other industrially used chemicals. Also, Khan and colleagues. (2019) used this descriptor to study interspecies modelling of *Cloeon dipterum* toxicity and *D. magna* toxicity and unsaturation index has highly contributed to the Endocrine Disrupting Chemical (EDC) toxicity against respective species.

HATS1s is the second important descriptor of the model, it belongs to the Geometry, Topology and Atom-Weight Assembly (GETAWAY) descriptors which are derived from Molecular Influence Matrix (MIM) (Todeschini and Consonni, 2008). HATS indices can be defined as weighting each atom of the molecule by its physico-chemical properties in combination with the diagonal elements of the molecular influence matrix. So, 3D properties of the molecules are also taken into account. MIM (denoted by H) can be defined in the following equation (Eq. 4.7):

$$H = M \times (M^T \times M)^{-1} \times M^T \tag{4.7}$$

where M is the molecular matrix consisted of Cartesian coordinates of the molecule atoms (hydrogens included) in a chosen structure. GETAWAY descriptors previously used by Filho and colleagues (2016) for modelling of larvicidal activity of monoterpenes and derivatives against *Aedes aegypti* and by Macêdo and colleagues. (2018) for modelling of antimalarial activity of dihydroartemisin and 19 derivatives.

H-046 has positively contributed to the larvicidal activity. It belongs to the atom centered fragment descriptors that helps to describe atoms with their own types, their bond types and their first neighbors. The atom-centered fragment descriptors are basic molecular descriptors that define the number of specific atom types in a molecule (i.e. H attached to $C^0(sp^3)$ no X attached to next C). Yangjeh and Jenagharad (2009) used this descriptor to calculate the toxicity of phenols to *Tetrahymena pyriformis* and found that the toxicity strongly related with this descriptor. Also, Yang and colleagues(2013) used this descriptor to investigate the QSAR of insecticidal activity of cholesterol-based hydrazine derivatives to third instar oriental armyworm (*Mythimna separata*). In addition, another atom centered fragment descriptor, H-052, was used for modelling insecticidal activity of plant derived compounds against *Aedes aegypti* (Saavedra et al., 2018).

GATS7e which is one of the 2D autocorrelations descriptors, has a significant effect in the model. It is a leverage weighted autocorrelation descriptor weighted by Sanderson electronegativity. Sanderson scale for electronegativity is based on covalent radii. GATS7e represents the electronegativity values of atoms separated by a topological distance of 7 bonds. 2D autocorrelation descriptors have been also used in previous studies for modelling of 33 isoxazoline and oxime derivatives of podophyllotoxin as insecticidal agents (Wang et al. 2012), and modelling of 55 monoterpenes against *Aedes aegypti* larvae (Santos et al., 2018).

Another descriptor in the model is Mor30s and it belongs to the 3D-MoRSE (Molecule Representation of Structures based on Electron diffraction) descriptors and depends on the three-dimensional structure of a molecule (Schuur et al., 1996). An equation was generated by Schuur and colleagues. (1996) using equations (Eq. 4.8) published previously for the atomic properties:

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i A_j \frac{sinsr_{ij}}{sr_{ij}} \qquad (4.8)$$

where I(s) is the intense of scattered radiation. The value of *s* ranges between 0−31.0 $Å^{-1}$ from the three-dimensional atomic coordinates of a molecule. Ai and Aj show properties of atoms at i and j positions and rij represents the interatomic distances. 3D-Morse descriptors were used by Duschowicz and colleagues (2009) for modelling the antifeedant activity of flavone derivatives.

The generated model is promising not only for being comparable to the literature (Table 4.10) but also for having the best activity range:

Table 4.10. Linear QSAR models on larvicidal activity from different studies.

Model	Chemical class	Number of compounds	Tr/Test Set division	Range (M)	Number of descriptors	R^2	Reference
1	Piperidine derivative	33	25/8	1.47	4	0.86	Doucet et al., 2017
2	Terpenes, phenylpropanoids	55	41/14	1.81	4	0.75	Carmenate et al., 2017
3	Monoterpenes and structurally related compounds	31	24/7	1.72	9	0.83	Filho et al., 2016
4	Terpenes, phenylpropanoids and oxygenated compounds	60	50/10	Range in ppm	5	0.84	Saavedra et al., 2017
Full model	Terpenes, phenylpropanoids, alkaloids, quinone derivatives	82	66/16	2.76	5	0.76	**Present study**

4.2. Prediction of Aquatic Toxicity of Vector Control Chemicals

4.2.1. 72-h algal toxicity model

The 72-h algal toxicity of the 230 chemicals (dataset and external set chemicals) was evaluated with using model (Eq. 4.9) reported by Önlü and Saçan (2017a). The model predicted the aquatic toxicity of 179 compounds out of 230. The toxicity of the 230 chemicals (complete dataset) was evaluated with the 72-h algal toxicity model in equation 4.9. The model predicted the aquatic toxicity of 179 compounds out of 230. The structural coverage of the model was 77.8%. The Insubria graph of the model is shown in Figure 4.5. The name of chemicals and predicted algal toxicity values from Eq. 4.9 together with their hat and descriptor values are provided in Appendix E (Table E1).

pEC_{50} = 5.140 (±0.473) + 3.484 (±0.710) SPAM + 1.924 (±0.251) Mor31p
+ 0.237 (±0.059) NdsCH – 0.439 (±0.093) CATS2D_02_AP + 0.950 (± 0.151) B05 [C-S]
+ 0.150(±0.015) F03 [C-N]+ (0.098 ±0.007) MLOGP2
-0.765 (±0.074) Hardness (4.9)

n=455; R^2= 0.661; Q^2_{LOO}= 0.643; CCC_{Tr}= 0.796; $RMSE_{Tr}$= 0.653; $RMSE_{CV}$= 0.670; MAE_{test}= 0.438

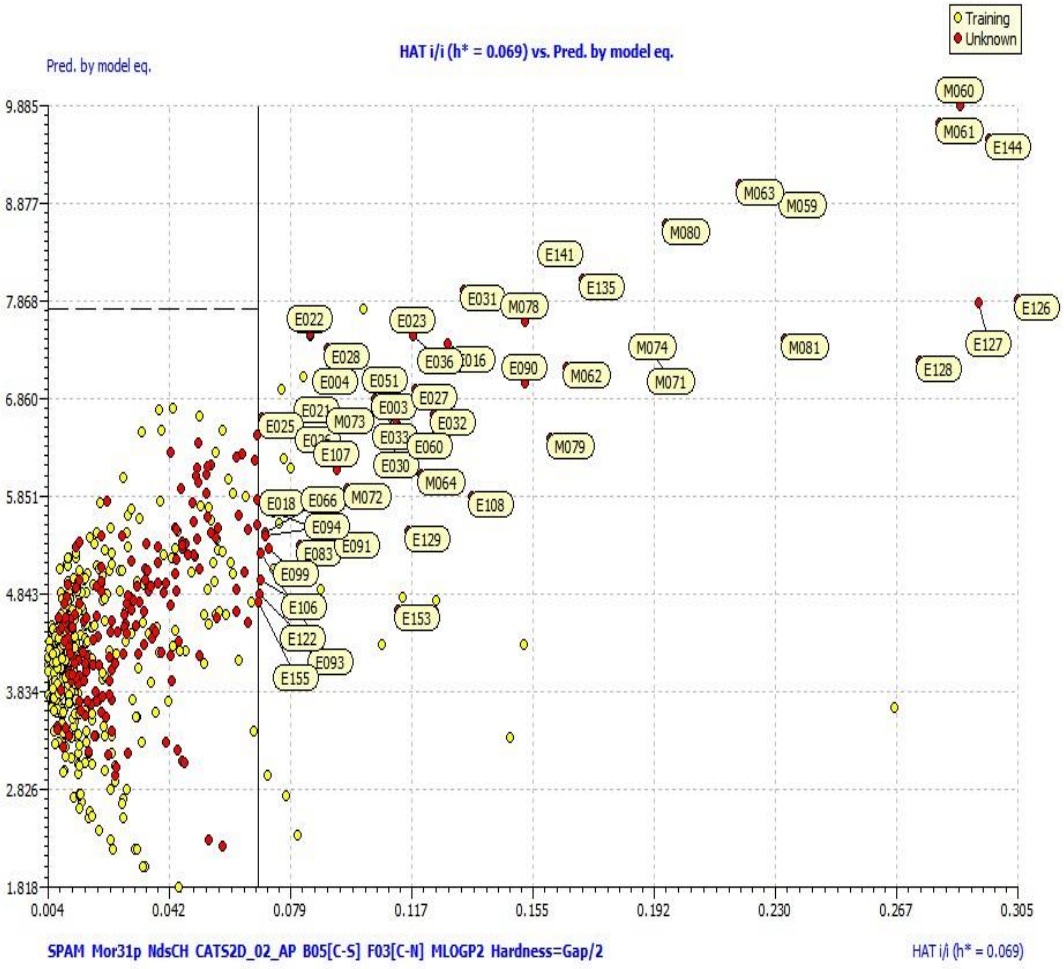

Figure 4. 5. Insubria graph of the algal toxicity model.

The most and the less toxic compounds screened from algal model is shown in Table 4.11. Descriptors appeared in 72h algal toxicity model and their meanings are given in Appendix F (Table F1).

Table 4.11. The most and the least toxic 10 chemicals for algae screened from the complete dataset using Eq. 4.9.

	Name	Predicted pEC_{50} value from Eq. 4.9 (mol/L)
The most toxic	epizonarene	6.41
	curcumene	6.30
	delta-cadinene	6.18
	valencene	6.17
	g-elemene	6.15
	beta-guaiene	6.08
	alpha-copaene	6.06
	caryophyllene	6.00
	alpha-santalene	5.93
	beta-selinene	5.89
The least toxic	isoborneol	2.24
	borneol	2.31
	p-menthane-3,8-diol	2.97
	1,8-cineole	3.06
	5-norbornene-2,2-dimethanol	3.11
	5-norbornene-2-ol	3.12
	1,4-cineole	3.19
	menthol	3.20
	3,4,5-trimethoxy toluene	3.23
	5-norbornene-2-endo-3-endo-dimethanol	3.24

4.2.2. Rainbow trout (*Oncorhynchus mykiss*) liver cell line RTL-W1 cytotoxicity model

The cytotoxicity of the 230 chemicals (complete dataset) was evaluated with model (Eq. 4.10) reported by Önlü and Saçan (2017b). The model predicted the cytotoxicity of 226 compounds out of 230. The other four compounds are within the structural applicability domain, but their predictions are not reliable. The Insubria graph of the cytotoxicity model is shown in Figure 4.6. The name of chemicals and predicted cytotoxicity values from Eq. 4.10 together with their hat and descriptor values are provided in Appendix E (Table E2).

pLC_{50} = + 2.8585 (±2.0573) -1.0979 (±0.4018) nRCOOH
+ 0.4529 (±0.2186) E_{HOMO} (eV) (4.10)

n=13; R^2= 0.839; Q^2_{LOO}= 0.727; CCC_{Tr}= 0.912; $RMSE_{Tr}$= 0.261; $RMSE_{CV}$= 0.339

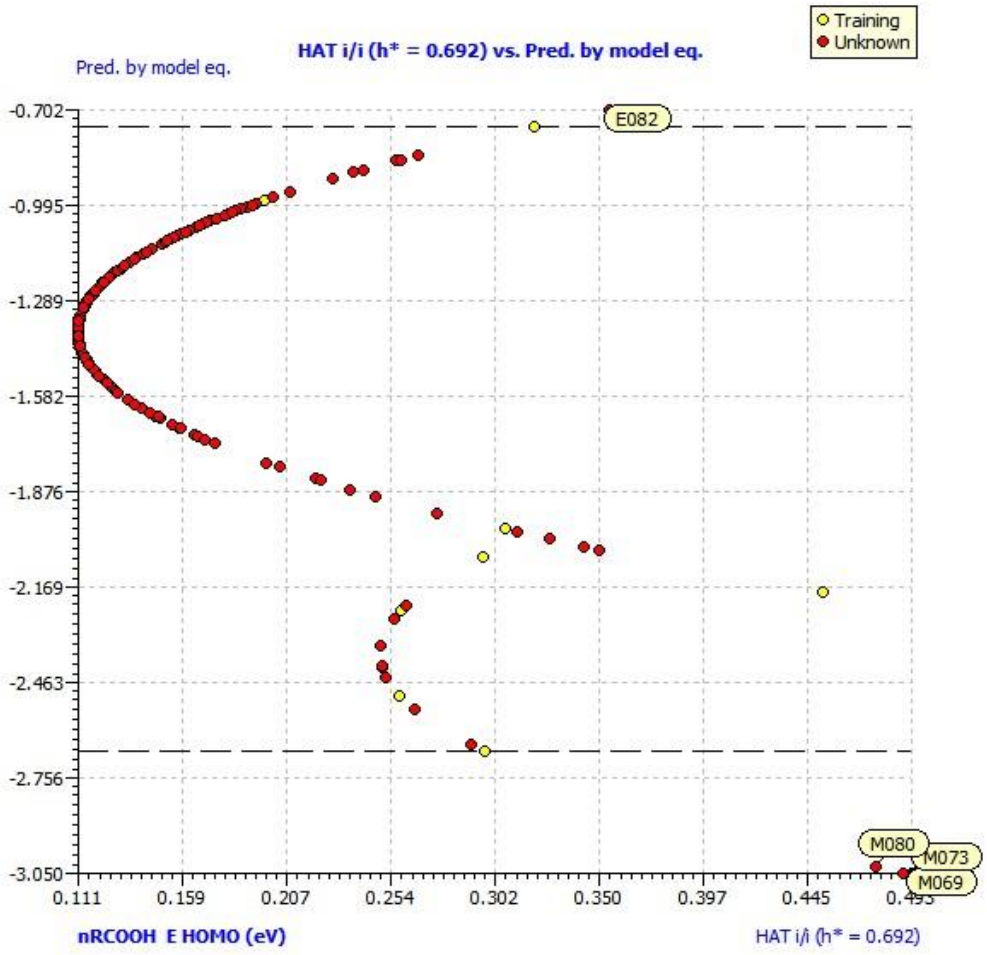

Figure 4.6. Insubria graph of RTL-W1 cytotoxicity model.

Descriptors appeared in the model with their meanings shown in Appendix F (Table F2). The most and the least cytotoxic compounds screened from Eq. 4.10 are listed in Table 4.12.

Table 4.12. The most and the least toxic 10 chemicals for RTL-W1 screened from the complete dataset using Eq. 4.10.

	Name	Predicted pLC_{50} value from Eq. 4.10 (µmol/L)
Most toxic	epi-zonarene	-0.83
	Z-asarone	-0.83
	1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	-0.85
	asaricin	-0.85
	1,2-dimethoxy-4-(2-propen-1-yl)-benzene	-0.88
	apiole	-0.89
	alpha-terpinene	-0.91
	guineensine	-0.95
	trans-anethole	-0.95
	locustol	-0.96
Less toxic	cinnamic acid	-2.65
	linoleic acid	-2.54
	geranic acid	-2.54
	2-(2-methoxy-4-(2-propen-1-yl))-phenoxy acetic acid	-2.44
	citronellic acid	-2.41
	carvacryl glycolic acid	-2.40
	p-methoxy cinnamic acid	-2.34
	ferulic acid	-2.26
	thymoxyacetic acid	-2.22
	hexyl butyrate	-2.05

4.2.3. *Dugesia japonica* model

The toxicity of the 141 chemicals out of 230 (complete dataset) was evaluated with model (Eq. 4.11) which is generated by Önlü and Saçan (2018), because this model is generated with the log K_{ow} values retrieved from Danish QSAR database and the log K_{ow} values of only 141 chemicals are available. The model predicted the aquatic toxicity of 133 compounds out of 141. The structural coverage of the model was 94.3%. Ten compounds are within the structural domain of the model, but the prediction is not reliable because they are not within the response domain. The Insubria graph of the model is shown below in Figure 4.7. The name of chemicals and predicted planarian toxicity values from Eq. 4.11 together with their hat and descriptor values are provided in Appendix E (Table E3).

$$pLC_{50} = -10.415\ (\pm 1.861) + 0.279\ (\pm 0.040)\ \log K_{ow} + 1.132\ (\pm 0.219)\ GATS7p + 6.604\ (\pm 1.865)\ SpMaxA\_G/D + 0.110\ (\pm 0.040)\ CATS2D\_08\_DL + 0.147\ (\pm 0.055)\ Mor31s \quad (4.11)$$

Descriptors appeared in Eq. 4.11 and their meanings are given in Appendix F (Table F3). The most and the least toxic compounds screened from external set using Eq. 4.11 is shown in Table 4.13.

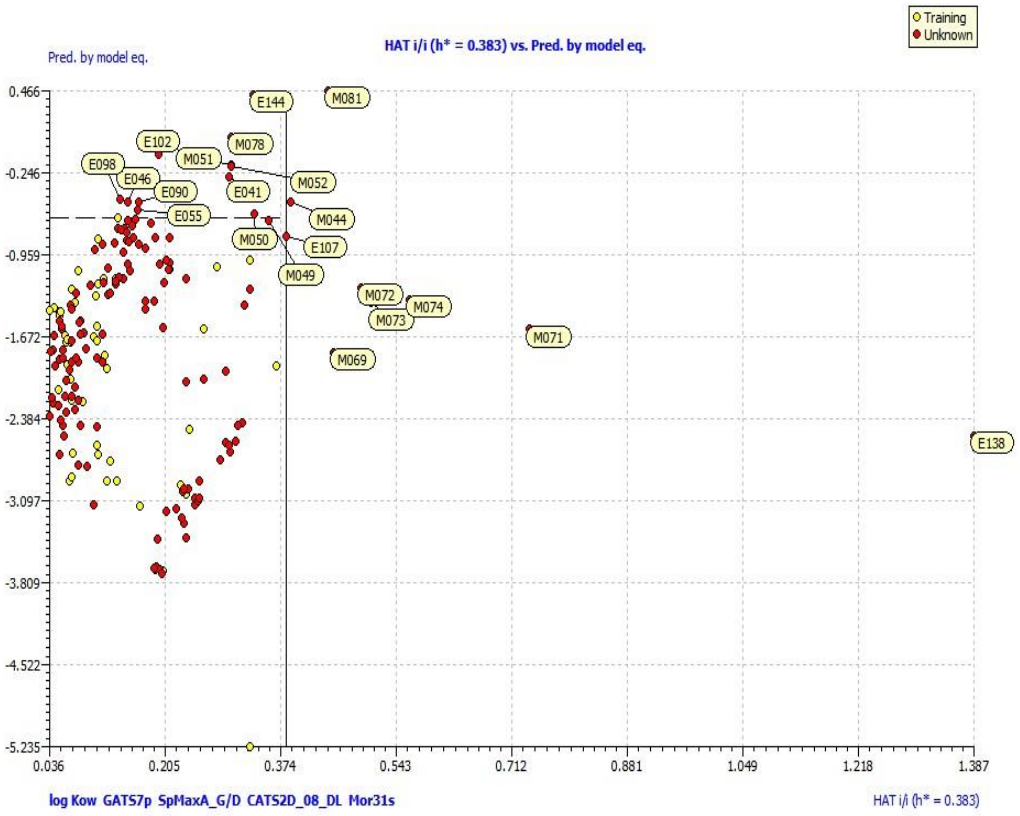

Figure 4.7. Insubria graph of *D. japonica* model.

The chemical “E138: rutin” is excluded from the predicted chemicals because the hat value of this compound is greater than h^*= 0.383. So, Insubria graph of *D. japonica* model is shown in Figure 4.8.

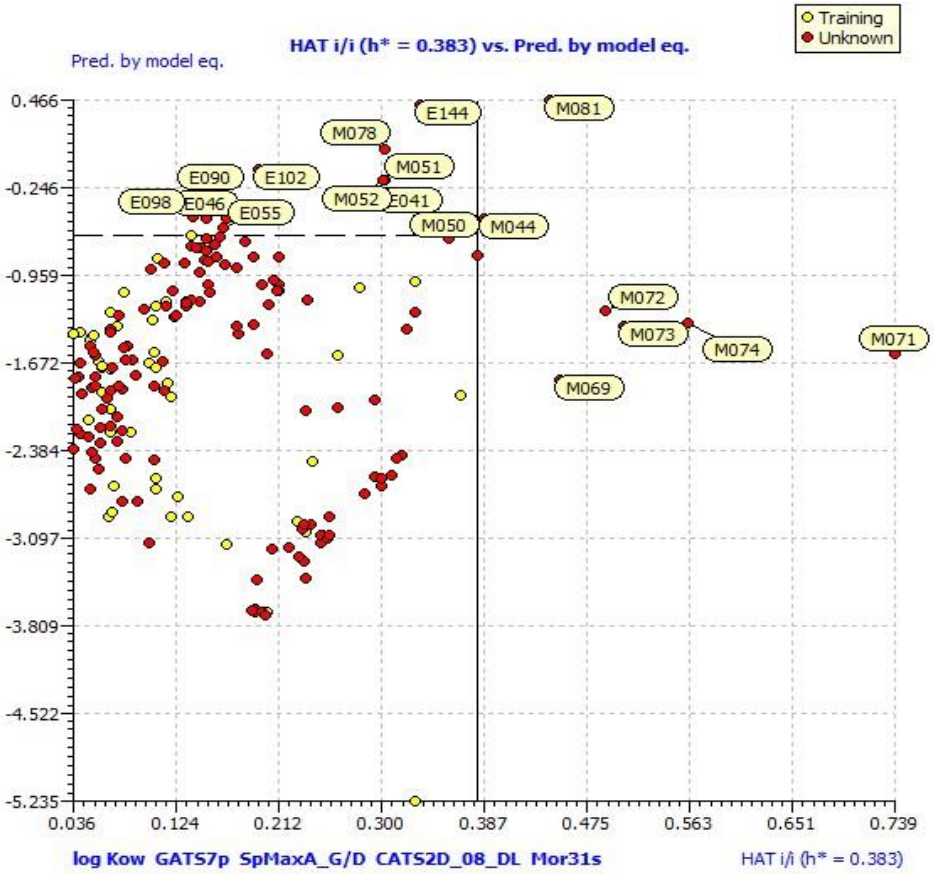

Figure 4.8. Insubria graph of *D. japonica* model (revised).

Table 4.13. The most and the least toxic 10 chemicals for *Dugesia japonica* screened from the complete dataset using Eq. 4.11.

	Name	Predicted pLC_{50} value from Eq. 4.11(μmol/L)
Most toxic	beta-selinene	-0.08
	S-limonene	-0.17
	delta-cadinene	-0.49
	beta-phellandrene	-0.64
	g-terpinene	-0.65
	(E)-beta-ocimene	-0.68
	1-dodecanol	-0.71
	beta-eudesmol	-0.72
	2,4-ditert-butylphenol	-0.73
	alpha-terpinene	-0.75
Less toxic	vanillin	-3.72
	catechol	-3.69
	guaiacol	-3.68
	resorcinol	-3.67
	5-norbornene-2-ol	-3.67
	phenol	-3.42
	para-benzoquinone	-3.41
	2-methyl-para-benzoquinone	-3.28
	borneol	-3.23
	salicylaldehyde	-3.18

4.3. Comparison of The Results for Finding Safe Larvicide

In order to propose environmentally safe larvicide, aquatic toxicity of larvicidal chemicals were predicted using previously published three aquatic toxicity models (algae, cytotoxicity and *Dugesia japonica*). There is no common larvicide that is safe for the three species as shown in Table 4.14 that represents the 10 least toxic compounds.

Table 4.14. Comparison of the least toxic chemicals for the three aquatic species.

The least toxic 10 chemicals[a]					
Algae	**pEC_{50} (mol/L)**	**RTL-W1**	**pLC_{50} (mol/L)**	***D. japonica***	**pLC_{50} (mol/L)**
p-menthane-3,8-diol	2.97	(E)-cinnamic acid	3.35	*para-benzoquinone*	2.59
3,4,5-trimethoxytoluene	3.23	citronellic acid	3.59	*2-methyl-para-benzoquinone*	2.72
quercetin	3.43	*p*-methoxycinnamic acid	3.66	*myrtenol*	2.91
*3,5-dimethoxytoluene**	3.48	*ferulic acid*	3.74	*fenchone*	2.99
coumestrol	3.57	hexyl butyrate	3.95	eucarvone	3.02
fenchone	3.59	octyl acetate	3.96	verbenone	3.02
emodic acid	3.63	*para-benzoquinone*	4.07	*3,5-dimethoxytoluene*	3.21
bornyl acetate	3.64	*2-methyl-para-benzoquinone*	4.12	*ferulic acid*	3.22
myrtenol	3.74	1-dodecanol	4.14	*cis*-isolongifolone	3.27
fenchene	3.75	2-isopropyl-para-benzoquinone	4.17	beta-pinene	3.34

[a]Based on the predicted values from Eq. 4.9 (algae), Eq. 4.10 (RTL-W1) and Eq. 4.11 (*D. japonica*). Toxicity values reported in the same unit for comparison *Common chemicals from each group are *italic*.

The common chemicals from each group were highlighted to compare their toxicities, and the least toxic chemical was proposed as a "safe" larvicide. The chemicals with their predicted toxicity values for each species are shown in Table 4.15:

Table 4.15. Comparison of the predicted aquatic toxicity of common larvicides.

Chemical	Algae (pEC_{50}) Eq. 4.9 (mol/L)	RTL-W1 (pLC_{50}) Eq. 4.10 (mol/L)	*D. japonica* (pLC_{50}) Eq. 4.11 (mol/L)
myrtenol	**3.74**	**4.59**	**2.91**
fenchone	3.59	4.64	2.99
3,5-dimethoxytoluene	3.48	4.91	3.21
ferulic acid	4.33	3.74	3.22
para-benzoquinone	out of AD	4.07	2.59
2-methyl-para-benzoquinone	4.56	4.12	2.72

The results show that "myrtenol" (Figure 4.9) can be environmentally safe larvicide, but according to the prediction made by model (Eq. 4.3) this compound cannot be considered as very effective considering its pLC_{50} (2.55).

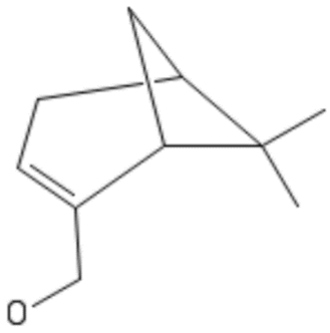

Figure 4.9. Chemical structure of myrtenol (The structure was drawn using PubChem Sketcher v.2.4).

Among the least toxic chemicals for three representative species, the common feature is that all of the compounds contain at least one oxygen atom attached to the ring. As an oxygenated monoterpene "myrtenol" was found as environmentally safe larvicide and its chemical and structural parameters like lipophilicity, polarizability and the energy (E_{HOMO} and Hardness) effect the aquatic toxicity of this compound to three representative species. Nevertheless, the compound cannot be predicted as an effective larvicide due to its low unsaturated bonds (Ui=1) and other parameters like electronegativity, 3D structure.

4.4. Comparison of the Results for Finding Effective and Safe Larvicide

The most toxic 10 chemicals predicted by model (Eq. 4.3) were screened with aquatic toxicity models for three representative species (algae, fish (RTL-W1) and *D. japonica*) in order to propose both environmentally safe and effective larvicide as listed in Table 4.16:

Table 4.16. Comparison of the predicted aquatic toxicity values** of chemicals with the highest larvicidal activity.

Chemical	***Aedes aegypti* larvae (pLC_{50}) Eq. 4.3**	**Algae (pEC_{50}) Eq. 4.9**	**RTL-W1 (pLC_{50}) Eq. 4.10**	***D. japonica* (pLC_{50}) Eq. 4.11**
1-undec-10-enoyl-4-benzyl-piperidine	4.34	out of AD	4.70	-*
pipercide	4.28	out of AD	5.05	-
1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	4.17	out of AD	4.70	-
3-benzyl-1-undec-10-enoyl-piperidine	4.17	out of AD	4.72	-
retrofractamide a	4.13	out of AD	4.94	-
2-benzyl-1-undec-10-enoyl-piperidine	4.07	out of AD	4.71	-
1-octanoyl-3-benzyl piperidine	4.01	out of AD	4.72	-
alpha-bisabolol	4.00	5.48	4.74	4.93
tectoquinone	**3.96**	**4.61**	**4.37**	**5.11**
alpha-eudesmol	3.96	5.38	4.84	5.30

*Cannot be predicted because of missing logK_{ow} values. **Toxicity values reported in the same unit (mol/L) for comparison.

Among the most toxic chemicals for *Aedes aegypti*, "tectoquinone" as shown in Figure 4.10 was highlighted as relatively both effective and safe larvicide. The predicted high larvicidal activity of tectoquinone is likely due to its high unsaturation index (3.17) which is the most important descriptor appeared in the generated model (Eq. 4.3). It is also toxic for the three aquatic species particularly due to its lipophilicity (logK_{ow}=3.89).

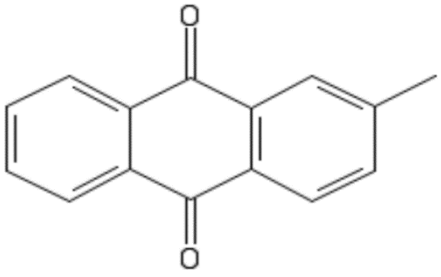

Figure 4.10. Chemical structure of tectoquinone (The structure was drawn using PubChem Sketcher v.2.4).

Nevertheless, toxicity values of most of the chemicals for *D. japonica* cannot be predicted because of missing logK_{ow} values, that's why the predicton is limited about consistency. Also, most of the chemicals are out of applicability domain of algae model. That's why it is hard to propose both environmentally safe and effective larvicide regarding the results.

6. CONCLUSIONS

In the present study, a QSAR model was generated for the prediction of larvicidal activity of plant-based compounds against 3rd and early 4th instar larvae of *Aedes aegypti* which is the main vector of Zika virus. The model was validated both internally and externally according to the requirements of Organization of Economic Co-operation Development (OECD) principles.

The model was generated by using Multiple Linear Regression (MLR) based on Ordinary Least Square (OLS) method of QSARINS software. Various training and test set divisions of 82 structurally diverse chemicals led to generate many QSAR models. Of the generated models, 5-descriptor models were evaluated in terms of *MAE*-based criteria and their predictive ability. Four over ten models were found to be in line with *MAE*-based criteria. The generated 5 descriptor models were externally tested with 148 chemicals mainly comprising piperidine derivatives, monoterpenoids, sesquiterpenoids and phenylpropanoids. The potential of these chemicals to be an effective larvicide has been searched, since they have neither experimental nor predicted larvicidal activity data reported in the literature. Although the external data set is very diverse the structural coverage of these four models were over 80%. After the assessment of the predictivity of these models, the prediction/test set is re-included in the training set and a full model comprising all the available information was computed. As a consequence of structural heterogeneity of the dataset, the generated linear QSAR model included complex and different descriptors. Four of the molecular descriptors appeared in the proposed model were derived from two-dimensional structure and one (Mor30s) was derived from three-dimensional structure of each chemical. The most important descriptor, namely unsaturation index (Ui), reveals that the number of SO_3 groups and the abundance of unsaturated bonds in the structure of a chemical increase the larvicidal activity. It is likely that as the unsaturation increases, the lipophilic character increases as well and it makes easier to go through larvae cuticle, thus larvicidal activity increases with unsaturation. Based on QSAR model results, the geometry, topology and atoms weighted by physicochemical properties (HATS1s), atom-centered fragments with specific atom types in a molecule (H-046) and electronegativity-based parameter indicated by GATS7e were found to be important factors for larvicidal activity.

The full model has 95.3% structural coverage for the chemicals in the external dataset. The predicted larvicidal activity values of external set chemicals which fell in the AD of model were used to screen the top ten chemicals with the most and the least larvicidal activity. Among the

studied plant-based compounds, piperidine derivatives are highlighted as the highest larvicidal activity chemicals for *Aedes aegypti*. Of the 10 compounds with highest larvicidal activity, there are 5 piperidine derivatives.

Besides fulfilling the data gap in the literature with the predicted larvicidal activity of plant-based larvicides in a way of having broader range for pLC_{50} value (more than 2.0 log unit) which is a requirement for a valid QSAR model, another contribution of the present study is the prediction of aquatic toxicity of plant-based larvicides for the three aquatic species (algae, fish (RTL-W1), and planarian). Based on the predicted toxicity values, the least toxic chemicals for algae were found as structures with -OH substitution, namely alcohols. This situation can be interpreted in the model as increasing toxicity with increasing hydrophobicity. The least toxic compounds for RTL-W1 were found as structures with -COOH substitution, namely carboxylic acids. This situation can be expressed as the model itself has a negative sign for the number of aliphatic carboxylic acid descriptor (nRCOOH). Moreover, the least toxic chemicals for *D. japonica* were found as structures with low $\log K_{ow}$ values. It is reasonable that the toxic action can be due to the lipophilicity of the compounds for the perturbation of membrane function.

In the present study, although the generated QSAR model for larvicidal activity proposes piperidine derivatives as potential larvicide, the algal toxicity of most of the piperidine derivatives couldn't be calculated. As such, the predicted algal toxicity values of these chemicals were mostly out of the applicability domain of the algal model. Nevertheless, they were found moderately toxic for RTL-W1. One of the purposes of this study was to find a plant-based larvicide that is effective on *Aedes aegypti* larvae. Regarding the larvicidal activities of several extracts of *Piper nigrum* fruit, it was proposed as a source for an effective larvicide. We also aimed to propose an environmentally safe larvicide in the present study. In this sight, the compound that is safe for three representative species was predicted as "myrtenol", but this compound is not very effective as a larvicide regarding its pLC_{50} value (2.55 (mol/L)). This compound can also be found in *Piper nigrum*. On the other hand, "tectoquinone" was proposed as both effective and relatively environmentally safe larvicide by comparing the aquatic toxicity values of most effective larvicides predicted by the proposed model for three representative species.

Finally, this study might suggest that *Piper nigrum* should be further investigated the possibility of finding both effective and environmentally friendly larvicide. A mixture of plant-based extracts can also be considered as larvicide regarding their possible synergistic effect. This

finding is very beneficial and has a potential to bridge the green chemistry and sustainability in the environment.

REFERENCES

Abe, F.R., Coleone, A.C., Machado, A.A., Machado-Neto, J.G., 2014. Ecotoxicity and Environmental Risk Assessment of Larvicides Used in the Control of *Aedes aegypti* to *Daphnia magna* (Crustacea, Cladocera). Journal of Toxicology and Environmental Health, Part A, 77, 37-45.

Albuquerque, M.R.J.R., Silveira, E.R., De A. Uchôa, D.E., Lemos, T.L.G., Souza, E.B., Santiago, G.M.P., Pessoa, O.D.L., 2004. Chemical composition and larvicidal activity of the essential oils from Eupatorium betonicaeforme (DC) Baker (Asteraceae). Journal of Agricultural and Food Chemistry, 52, 6708-6711.

Ali, A., Murphy, C.C., Demirci, B., Wedge, D.E., Sampson, B.J., Khan, I.A., Baser, H.C., Tabanca, N., 2013. Insecticidal and biting deterrent activity of rose-scented geranium (Pelargonium spp.) essential oils and individual compounds against *Stephanitis pyrioides* and *Aedes aegypti*. Pest Management Science, 69, 1385-1392.

Cantrell, C.L., Pridgeon, J.W., Fronczek, F.R., Becnel, J.J., 2010. Structure–activity relationship studies on derivatives of Eudesmanolides from *Inula helenium* as toxicants against *Aedes aegypti* larvae and adults. Chemistry & Biodiversity, 7, 1681–1697.

Canizares-Carmenate, Y., Hernandez-Morfa, M., Torrens, F., Castellano, G., Castillo-Garit, J.A., 2017. Larvicidal activity prediction against *Aedes aegypti* mosquito using computational tools. Journal of Vector Borne Diseases, 54, 164.

Cheng, S.S., Liu, J.Y., Tsai, K.H., Chen, W.J., Chang, S.T., 2004. Chemical composition and mosquito larvicidal activity of essential oils from leaves of different *Cinnamomum osmophloeum* provenances. Journal of Agricultural Food Chemistry, 52, 4395-4400.

Cheng, S.S., Huang, C.G., Chen, Y.J., Yu, J.J., Chen, W.J., Chang, S.T., 2009. Chemical compositions and larvicidal activities of leaf essential oils from two *Eucalyptus* species. Bioresource Technology, 100, 452-456.

Cheng, S.S., Lin, C.Y., Chung, M.J., Liu, Y.H., Huang, C.G., Chang, S.T., 2013. Larvicidal activities of wood and leaf essential oils and ethanolic extracts from *Cunninghamia konishii* Hayata against the dengue mosquitoes. Industrial Crops and Products, 47, 310-315.

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.

Chizzola R., 2013. Regular Monoterpenes and Sesquiterpenes (Essential Oils). In: Ramawat K., Mérillon J.M. (Eds), Natural Products, 2973-3008, Springer, Berlin, Heidelberg.

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

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

Costa, J.G.M., Rodrigues, F.F.G., Angélico, E.C., Silva, M.R., 2005 Chemical-biological study of the essential oils of *Hyptis martiusii*, *Lippia sidoides* and *Syzigium aromaticum* against larvae of *Aedes aegypti* and *Culex quinquefasciatus*. Revista Brasileira de Farmacognosia,15, 304–309.

Costa, J.G.M., Rodrigues, F.F.G., Sousa, E.O., Junior, D.M.S., Campos, A.R., Coutinho, H.D.M., Lima, S.G., 2010. Composition and larvicidal activity of the essential oils of *Lantana camara* and *Lantana montevidensis*. Chemistry of Natural Compounds, 46, 313–315.

De, P., Roy, K., 2018. Greener chemicals for the future: QSAR modelling of the PBT index using ETA descriptors. SAR and QSAR in Environmental Research, 29, 319-337.

Devillers, J., Lagneau, C., Lattes, A., Garrigues, J.C., Clémenté, M.M., Yébakima, A., 2014. In silico models for predicting vector control chemicals targeting *Aedes aegypti*. SAR and QSAR in Environmental Research, 25, 803–835.

Devillers, J., Doucet-Panaye, A., Doucet, J., 2015. Structure–activity relationship (SAR) modelling of mosquito larvicides. SAR and QSAR in Environmental Research, 26, 263-278.

Dias C.N., Moraes D.F.C., 2014. Essential oils and their compounds as *Aedes aegypti* L. (Diptera: Culicidae) larvicides: review. Parasitology Research, 113, 565-592.

Doucet, J.P., Papa, E., Doucet-Panaye, A., Devillers, J., 2017. QSAR models for predicting the toxicity of piperidine derivatives against *Aedes aegypti*. SAR and QSAR in Environmental Research, 28, 451-470.

Duchowicz, P.R., Goodarzi, M., Ocsachoque, M.A., Romanelli, G.P., Ortiz, E.D., Autino, J.C., Bennardi, D., Ruiz, D., Castro, E.A., 2009. QSAR analysis on *Spodoptera litura* antifeedant activities for flavone derivatives. Science of The Total Environment, 408, 277-285.

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.

Federici, B.A., 2003. Recombinant bacteria for mosquito control. Journal of Experimental Biology, 206 , 3877-3885.

Filho, E.B., Silva, J.W., Cavalcanti, S.C., 2016. Quantitative structure-toxicity relationships and molecular highlights about *Aedes aegypti* larvicidal activity of monoterpenes and related compounds. Medicinal Chemistry Research, 25, 2171-2178.

Friedrich H., 1976. Phenylpropanoid constituents of essential oils. Lloydia, 39, 1–7.

Goellner, E., Schmitt, A.T., Couto, J.L., Müller, N.D., Pilz-Junior, H.L., Schrekker, H.S., Silva, C., Silva, O.S., 2018. Larvicidal and residual activity of imidazolium salts against *Aedes aegypti* (Diptera: Culicidae). Pest Management Science,74 , 1013-1019.

Goindin, D., Delannay, C., Gelasse, A., Ramdini, C., Gaude, T., Faucon, F., Gustave, J., Vega-Rua, A., Fouque, F., 2017. Levels of insecticide resistance to deltamethrin, malathion, and temephos, and

associated mechanisms in Aedes aegypti mosquitoes from the Guadeloupe and Saint Martin islands (French West Indies). Infectious Diseases of Poverty, 6, 38.

Golbraikh, A., Tropsha, A., 2002. Beware of q2!. Journal of Molecular Graphics and Modelling, 20, 269-276.

Goldberg, D.E., Holland, J.H., 1988. Genetic algorithms and machine learning. Machine Learning, 3, 95-99.

Govindarajan, M., 2010. Chemical composition and larvicidal activity of leaf essential oil from Clausena anisata (Willd.) Hook. f. ex Benth (Rutaceae) against three mosquito species. Asian Pacific Journal of Tropical Medicine, 3, 874-877.

Govindarajan, M., Benelli G., 2016a. Eco-friendly larvicides from Indian plants: effectiveness of lavandulyl acetate and bicyclogermacrene on malaria, dengue and Japanese encephalitis mosquito vectors. Ecotoxicology and Environmental Safety, 133, 395–402.

Govindarajan, M., Benelli, G., 2016b. *Artemisia absinthium*-borne compounds as novel larvicides: Effectiveness against six mosquito vectors and acute toxicity on non-target aquatic organisms. Parasitology Research, 115, 4649-4661.

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

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

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

Gulzar, T., Uddin, N., Siddiqui, B.S., Naqvi, S.N., Begum, S., Tariq, R.M., 2013. New constituents from the dried fruit of *Piper nigrum* Linn., and their larvicidal potential against the Dengue vector mosquito *Aedes aegypti*. Phytochemistry Letters, 6, 219-223.

Guyton, K.Z., Loomis, D., Grosse, Y., Ghissassi, F.E., Benbrahim-Tallaa, L., Guha, N., Straif, K., 2015. Carcinogenicity of tetrachlorvinphos, parathion, malathion, diazinon, and glyphosate. The Lancet Oncology, 16, 490-491.

Habibi-Yangjeh, A., Danandeh-Jenagharad, M., 2009. Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis. Monatshefte Für Chemie - Chemical Monthly, 140, 1279-1288.

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.

Hansch, C., Verma, R. P., 2009. Larvicidal activities of some organotin compounds on mosquito larvae: A QSAR study. European Journal of Medicinal Chemistry, 44, 260-273.

Hematpoor, A., Liew, S.Y., Chong, W.L., Azirun, M.S., Lee, V.S., Awang, K., 2016. Inhibition and Larvicidal Activity of Phenylpropanoids from *Piper sarmentosum* on Acetylcholinesterase against Mosquito Vectors and Their Binding Mode of Interaction. Plos One, 11, e0155215.

Jantan, I., Yalvema, M.F., Ahmad, N.W., Jamal, J.A., 2005. Insecticidal activities of the leaf oils of eight *Cinnamomum* species against *Aedes aegypti* and *Aedes albopictus*. Pharmaceutical Biology, 43, 526-532.

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

Khan, K., Roy, K., Benfenati, E., 2019. Ecotoxicological QSAR modeling of endocrine disruptor chemicals. Journal of Hazardous Materials, 369, 707-718. doi:10.1016/j.jhazmat.2019.02.019

Kode srl, Dragon (software for molecular descriptor calculation) version 7.0.10, 2017, https://chm.kode-solutions.net. Date accessed January 2019.

Kulkarni, R.R., Pawar, P.V., Joseph, M.P., Akulwad, A.K., Sen, A., Joshi, S.P., 2013. *Lavandula gibsoni* and *Plectranthus mollis* essential oils: chemical analysis and insect control activities against

Aedes aegypti, *Anopheles stephensi* and *Culex quinquefasciatus*. Journal of Pest Science, 86, 713-718.

Kweka, E.J., Lima, T.C., Marciale, C.M., Sousa, D.P., 2016. Larvicidal efficacy of monoterpenes against the larvae of *Anopheles gambiae*. Asian Pacific Journal of Tropical Biomedicine, 6, 290-294.

Lagadic, L., Roucaute, M., Caquet, T., 2013. Btisprays do not adversely affect non-target aquatic invertebrates in French Atlantic coastal wetlands. Journal of Applied Ecology, 51, 102-113.

Lima, G.P.G., Souza, T.M., Freire, G.P., Farias, D.F., Cunha, A.P., Ricardo, N.M.P.S., Morais, S.M, Carvalho A.F.U., 2013 Further insecticidal activities of essential oils from *Lippiasidoides* and *Croton* species against *Aedes aegypti* L. Parasitology Research, 112, 1953-1958.

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 C.H., Mishra A.K., Tan R.X., Tang C., Yang H., Shen Y.F., 2006. Repellent and insecticidal activities of essential oils from *Artemisia princeps* and *Cinnamomum camphora* and their effect on seed germination of wheat and broad bean. Bioresource Technology, 97, 1969-1973.

Lucia, A., Licastro, S., Zerba, E., Masuh, H., 2008. Yield, chemical composition, and bioactivity of essential oils from 12 species of *Eucalyptus* on *Aedes aegypti* larvae. Entomologia Experimentalis et Applicata, 129, 107-114.

Lucia, A., Juan, L.W., Zerba, E.N., Harrand, L., Marcó, M., Masuh, H.M., 2012. Validation of models to estimate the fumigant and larvicidal activity of *Eucalyptus* essential oils against *Aedes aegypti* (Diptera: Culicidae). Parasitology Research, 110, 1675-1686.

Lomonaco, D., Santiago, G.M., Ferreira, Y.S., Arriaga, Â.M., Mazzetto, S.E., Mele, G., Vasapollo, G., 2009. Study of technical CNSL and its main components as new green larvicides. Green Chemistry,11, 31-33.

Macêdo, W.J., Costa, J.S., Federico, L.B., Cruz, J.V., Carvalho, S.S., Ramos, R.S., Wanderley, D., Silva, C., Santos, C.B., 2018. A MLR and ADME/Tox Study of New Dihydroartemisinin Compounds with Antimalarial Activity. Journal of Computational and Theoretical Nanoscience,15, 1785-1794.

Magalhães, L.A.M., Lima, M.P., Marques, M.O.M., Facanali, R., Pinto, A.C.S., Tadei, W.P., 2010. Chemical composition and larvicidal activity against *Aedes aegypti* larvae of essential oils from four *Guarea* species. Molecules, 15, 5734-5741.

Marques, M.M.M., Morais, S.M., Vieira, Í.G.P., Vieira, M.G.S., Silva, A.R.A., Almeida, R.R., Guedes, M.I.F., 2011. Larvicidal activity of *Tagetes erecta* against *Aedes aegypti*. Journal of the American Mosquito Control Association, 27, 156-158.

Marques, A., Kaplan, M.A., 2014. Active metabolites of the genus Piper against *Aedes aegypti*: Natural alternative sources for dengue vector control. Universitas Scientiarum, 20, 61.

Melo-Santos, M., Varjal-Melo, J., Araújo, A., Gomes, T., Paiva, M., Regis, L., Furtado, T., Magalhaes, M., Macoris, M., Andrighetti, M., Ayres, C., 2010. Resistance to the organophosphate temephos: Mechanisms, evolution and reversion in an *Aedes aegypti* laboratory strain from Brazil. Acta Tropica, 113, 180-189.

Merritt, R.W., Lessard, J.L., Wessell, K.J., Hernandez, O., Berg, M.B., Wallace, J.R., Novak, J., Ryan, J., Merritt, B.W., 2005. Lack of Effects of Bacillus Sphaericus (Vectolex®) On Nontarget Organisms In A Mosquito-Control Program In Southeastern Wisconsin: A 3-Year Study. Journal of the American Mosquito Control Association, 21, 201-212.

Nascimento, J.C., David, J.M., Barbosa, L.C.A., Paula, V.F., Demuner, A.J., David, J.P., Conserva, L.M., Ferreira-Jr, J.C., Guimarães, E.F., 2013. Larvicidal activities and chemical composition of essential oils from *Piper klotzschianum* (Kunth) C DC. (Piperaceae). Pest Management Science, 69, 1267-1271.

Nelson M.J., 1986. *Aedes aegypti*: Biology and Ecology. Pan American Health Organization, Washington, D.C.

Nunes, R.K., Martins, U.N., Brito, T.B., Nepel, A., Costa, E.V., Barison, A., Santos, R.L., Cavalcanti, S.C., 2018. Evaluation of (–)-borneol derivatives against the Zika vector, *Aedes aegypti* and a non-target species, Artemia sp. Environmental Science and Pollution Research, 25, 31165-31174.

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.

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

Ojima, I., Iula, D.M., 1999. New Approaches to the Syntheses of Piperidine, Izidine, and Quinazoline Alkaloids by Means of Transition Metal Catalyzed Carbonylations (Vol. 13, pp. 371-412). Oxford, UK: Elsevier.

Önlü, S., Saçan, M.T., 2017a. 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., 2017b. 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.

Pandey, S.K., Tandon, S., Ahmad, A., Singh, A.K., Tripathi, A.K., 2013. Structure-activity relationships of monoterpenes and acetyl derivatives against *Aedes aegypti* (Diptera: Culicidae) larvae. Pest Management Science, 69, 1235-1238.

Paris M., David J., Despres L., 2011. Fitness costs of resistance to Bti toxins in the dengue vector *Aedes aegypti*. Ecotoxicology, 20, 1184-1194.

Park, I., Lee, S., Shin, S., Park, J., Ahn, Y., 2002. Larvicidal Activity of Isobutylamides Identified in *Piper nigrum* Fruits against Three Mosquito Species. Journal of Agricultural and Food Chemistry, 50, 1866-1870.

Pavela R., Benelli G., 2016. Essential oils as ecofriendly biopesticides? Challenges and constraints. Trends in Plant Science, 21, 1000–1007.

Pavela, R., Govindarajan, M., 2016. The essential oil from *Zanthoxylum monophyllum* a potential mosquito larvicide with low toxicity to the non-target fish *Gambusia affinis*. Journal of Pest Science, 90, 369-378.

Perumalsamy, H., Kim, N.J., Ahn, Y.J., 2009. Larvicidal activity of compounds isolated from Asarum heterotropoides against Culex pipiens pallens, *Aedes aegypti* and *Ochlerotatus togoi* (Diptera: Culicidae). Journal of Asia-Pacific Entomology, 46, 1420–1423.

Pinto, A.C.S., Nogueira, K.L., Chaves, F.C.M., da Silva, L.V.S., Tadei, W.P., Pohlit, A.M., 2012. Adulticidal activity of dillapiol and semi-synthetic derivatives of dillapiol against adults of Aedes aegypti (L.)(Culicidae). Embrapa Amazônia Ocidental-Artigo em periódico indexado (ALICE).

Pohlit, A., Rezende, A., Baldin, E.L., Lopes, N., Neto, V.D., 2011. Plant Extracts, Isolated Phytochemicals, and Plant-Derived Agents Which are Lethal to Arthropod Vectors of Human Tropical Diseases - A Review. Planta Medica, 77, 618-630.

Pridgeon, J.W., Meepagala, K.M., Becnel, J.J., Clark, G.G., Pereira, R.M., Linthicum K.J., 2007. Structure–Activity Relationships of 33 Piperidines as Toxicants Against Female Adults of *Aedes aegypti* (Diptera: Culicidae). Journal of Medical Entomology, 44, 263-269.

PubChem Sketcher v.2.4, https://pubchem.ncbi.nlm.nih.gov/edit2/index.html. Date accessed June 2019.

Puzyn, T., Leszczynski, J., Cronin, M. T., 2010. Recent advances in QSAR studies methods and applications. New York: Springer.

Rocha, D.K., Matos, O., Novo, M. T., Figueiredo, A.C., Delgado, M., Moiteiro, C., 2015. Larvicidal Activity against *Aedes aegypti* of Foeniculum Vulgare Essential Oils from Portugal and Cape Verde. Natural Product Communications, 10,1934578x1501000438.

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

Ruiz, C., Cachay, M., Domínguez, M., Velásquez, C., Espinoza, G., Ventosilla, P., Rojas, R., 2011. Chemical composition, antioxidant and mosquito larvicidal activities of essential oils from *Tagetes filifolia*, *Tagetes minuta* and *Tagetes elliptica* from Perú. Planta Medica,77, PE30.

Rücker, C., Rücker, G., Meringer, M., 2007. Y-Randomization and its variants in QSPR/QSAR Journal of Chemical Information and Modeling, 7, 2345–2357.

Saavedra, L.M., Romanelli, G.P., Rozo, C.E., Duchowicz, P.R., 2018. The quantitative structure–insecticidal activity relationships from plant derived compounds against chikungunya and zika *Aedes aegypti* (Diptera:Culicidae) vector. Science of The Total Environment, 610, 937-943.

Santos, R.P., Nunes, E.P., Nascimento, R.F., Santiago, G.M.P., Menezes, G.H.A., Silveira, E.R., Pessoa, O.D.L., 2006. Chemical composition and larvicidal activity of the essential oils of *Cordia leucomalloides* and *Cordia curassavica* from the northeast of Brazil. Journal of the Brazilian Chemical Society, 17, 1027–1030.

Santos, H.S., Santiago, G.M.P., Oliveira, J.P.P., Arriaga, A.M.C., Marques, D.D., Lemos, T.L.G., 2007. Chemical composition and larvicidal activity against *Aedes aegypti* of essential oils from *Croton zehntneri*. Natural Product Communications, 2, 1233–1236

Santos, S.R., Silva, V.B., Melo, M.A., Barbosa, J.D., Santos, R.L., Sousa, D.P., Cavalcanti, S.C., 2010. Toxic Effects on and Structure-Toxicity Relationships of Phenylpropanoids, Terpenes, and Related Compounds in *Aedes aegypti* Larvae. Vector-Borne and Zoonotic Diseases, 10, 1049-1054.

Santos, S.R.L., Melo, M.A., Cardoso, A.V., Santos, R.L.C., de Sousa, D. P., Cavalcanti, S.C.H., 2011. Structure-activity relationships of larvicidal monoterpenes and derivatives against *Aedes aegypti* Linn., Chemosphere., 84, 150-153.

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

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

Scotti, L., Scotti, M.T., Silva, V.B., Santos, S.R.L., Cavalcanti, S.C.H., Mendonça, F.J.B.Jr., 2014. Chemometric studies on potential larvicidal compounds against *Aedes aegypti*. Medicinal Chemistry, 10, 201-210.

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.

Shivakumar, M.S., Srinivasan, R., Natarajan, D., 2013. Larvicidal potential of some Indian medicinal plant extracts against *Aedes aegypti* (L.). Asian Journal of Pharmaceutical and Clinical Research., 6, 77-80.

Silva, W.J., Doria, G.A., Maia, R.T., Nunes, R.S., Carvalho, G.A., Blank, A.F., Alves P.B., Marçal, R.M., Cavalcanti S.C.H., 2008. Effects of essential oils on *Aedes aegypti* larvae: alternatives to environmentally safe insecticides. Bioresource Technolgies, 99, 3251–3255.

Sivakumar, R., Jebanesan, A., Govindarajan, M., Rajasekar, P., 2011. Larvicidal and repellent activity of tetradecanoic acid against *Aedes aegypti* (Linn.) and *Culex quinquefasciatus* (Say.) (Diptera: Culicidae). Asian Pacific Journal of Tropical Medicine, 4, 706-710

Sousa, D.P., Vieira, Y.W., Uliana, M.P., Melo, M.A., Brocksom, T.J., Cavalcanti, S.C., 2010. Larvicidal activity of para-Benzoquinones. Parasitology Research, 107, 741-745.

SPARTAN v. 16, 2017. Wavefunction, Inc., Irvine, CA, USA. http://wavefun.com/. Date accessed January 2019.

SPSS v. 25, 2017. IBM Corp, Armonk, NY. https://www.ibm.com/us-en/marketplace?lnk=mp. Date accessed January 2019

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.

Todeschini, R., Consonni, V., 2008. Handbook of Molecular Descriptors. 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.

Todeschini, R., Consonni, V., Mauri, A., Pavan, M., 2004. Detecting "bad" regression models: Multicriteria fitness functions in regression analysis. Analytica Chimica Acta, 515, 199-208.

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

Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C.P., Agrawal, R.K., 2011. Validation of QSAR models-strategies and importance. International Journal of Drug Design & Discovery, 3, 511-519.

Vogt, T., 2010. Phenylpropanoid Biosynthesis. Molecular Plant, 3, 2-20.

Waliwitiya, R., Kennedy, C.J., Lowenberger, C.A., 2009. Larvicidal and oviposition-altering activity of monoterpenoids, transanethole and rosemary oil to the yellow fever mosquito *Aedes aegypti* (Diptera: Culicidae). Pest Management Science, 65, 241-248.

Wang, Y., Shao, Y., Wang, Y., Fan, L., Yu, X., Zhi, X., Yang, C., Qu, H., Yao, X., Xu, H., 2012. Synthesis and Quantitative Structure–Activity Relationship (QSAR) Study of Novel Isoxazoline and Oxime Derivatives of Podophyllotoxin as Insecticidal Agents. Journal of Agricultural and Food Chemistry, 60, 8435-8443.

Wang, Z., Perumalsamy, H., Wang, M., Shu, S., Ahn, Y., 2015. Larvicidal activity of *Magnolia denudate* seed hydrodistillate constituents and related compounds and liquid formulations towards two susceptible and two wild mosquito species. Pest Management Science, 72, 897-906.

WHO, 2005. Report of the WHO Informal Consultation on the evaluation and testing of insecticides. Geneva, World Health Organization, (CTD/WHOPES/IC/2005.10).

WHO, 2017. Vector-borne diseases. Retrieved from https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases. Date accessed April 2018.

WHO, 2018. Zika Virus. Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/zika-virus. Date accessed April 2018.

WHO, 2019. Dengue and severe dengue. Retrieved from https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue. Date accessed June 2019.

Yang, C., Shao, Y., Zhi, X., Huan, Q., Yu, X., Yao, X., Xu, H. 2013. Semisynthesis and quantitative structure–activity relationship (QSAR) study of some cholesterol-based hydrazone derivatives as insecticidal agents. Bioorganic & Medicinal Chemistry Letters, 23, 4806-4812

APPENDIX A: EXTERNAL SET CHEMICALS FOR THE STUDY

Table A1. External set chemicals used in the study.

Label	Name	CAS Number	log k_{ow}	Molecular Weight	Reference
E001	g-elemene	3242-08-8	4.56	204.357	Cheng et al. (2009)
E002	1-(cyclohexylacetyl)-2-methyl-piperidine	NA	2.72	223.360	Pridgeon et al. (2007)
E003	(2R)-1-decanoyl-2-methyl-piperidine	NA	4.14	253.430	Pridgeon et al. (2007)
E004	1-dodecanoyl-2-methyl-piperidine	NA	4.97	281.484	Pridgeon et al. (2007)
E005	(2R)-1-heptanoyl-2-methyl-piperidine	NA	2.89	211.349	Pridgeon et al. (2007)
E006	1-(3-cyclohexylpropanoyl)-2-methyl-piperidine	NA	3.14	237.387	Pridgeon et al. (2007)
E007	1-[(4-methylcyclohexyl) carbonyl]-2-methyl-piperidine	NA	2.87	223.360	Pridgeon et al. (2007)
E008	(3S)-1-(1-methylcyclohexyl) carbonyl-3-methyl-piperidine	NA	3.27	237.387	Pridgeon et al. (2007)
E009	(3S)-1-(3-cyclohexylpropanoyl)-3-methyl-piperidine	NA	3.22	237.387	Pridgeon et al. (2007)
E010	(3S)-1-heptanoyl-3-methyl-piperidine	NA	2.97	211.349	Pridgeon et al. (2007)
E011	(3S)-1-(cyclohexylcarbonyl)-3-methyl-piperidine	NA	2.80	223.360	Pridgeon et al. (2007)
E012	1-decanoyl-4-methyl-piperidine	NA	4.15	253.430	Pridgeon et al. (2007)
E013	1-(4-cyclohexylbutanoyl)-4-methyl-piperidine	NA	3.57	251.414	Pridgeon et al. (2007)
E014	1-(cyclohexylcarbonyl)-4-methyl-piperidine	NA	2.55	209.333	Pridgeon et al. (2007)
E015	1-(3-cyclohexylpropanoyl)-4-methyl-piperidine	NA	3.15	237.387	Pridgeon et al. (2007)
E016	1-dodecanoyl-4-methyl-piperidine	NA	4.99	281.484	Pridgeon et al. (2007)
E017	1-(cyclohexylcarbonyl)-2-ethyl-piperidine	NA	3.03	223.360	Pridgeon et al. (2007)
E018	1-(3-cyclohexylpropanoyl)-2-ethyl-piperidine	NA	3.62	251.414	Pridgeon et al. (2007)
E019	1-propionyl-2-ethyl-piperidine	NA	1.71	169.268	Pridgeon et al. (2007)
E020	1-(3-cyclopentylpropanoyl)-2-ethyl-piperidine	NA	3.21	237.387	Pridgeon et al. (2007)
E021	1-nonanoyl-2-ethyl-piperidine	NA	4.21	253.430	Pridgeon et al. (2007)
E022	1-octanoyl-3-benzyl-piperidine	NA	4.99	301.474	Pridgeon et al. (2007)
E023	1-undec-10-enoyl-4-benzyl-piperidine	NA	5.91	341.539	Pridgeon et al. (2007)
E024	1-cyclohexylacetyl-4-benzyl-piperidine	NA	4.34	299.458	Pridgeon et al. (2007)
E025	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	NA	4.75	313.485	Pridgeon et al. (2007)
E026	2-methyl-1-undec-10-enoyl-piperidine	NA	4.29	265.441	Pridgeon et al. (2007)
E027	2-ethyl-1-undec-10-enoyl-piperidine	NA	4.77	279.468	Pridgeon et al. (2007)

Table A1. (Continued).

Label	Name	CAS Number	log k_{ow}	Molecular Weight	Reference
E028	2-benzyl-1-undec-10-enoyl-piperidine	NA	5.96	341.539	Pridgeon et al. (2007)
E030	3-ethyl-1-undec-10-enoyl-piperidine	NA	4.79	279.468	Pridgeon et al. (2007)
E031	3-benzyl-1-undec-10-enoyl-piperidine	NA	5.97	341.539	Pridgeon et al. (2007)
E032	4-methyl-1-undec-10-enoyl-piperidine	NA	4.30	265.441	Pridgeon et al. (2007)
E033	4-ethyl-1-undec-10-enoyl-piperidine	NA	4.72	279.468	Pridgeon et al. (2007)
E035	alpha-pinene	7785-26-4	2.90	136.238	Santos et al. (2006)
E037	safrole	94-59-7	0.37	162.188	Jantan et al. (2005)
E038	piperitone	89-81-6	3.07	152.237	Marques et al. (2011)
E040	trans-ocimenone	33746-72-4	3.15	150.221	Ruiz et al. (2011)
E041	terpinolene	586-62-9	2.81	136.238	Cheng et al. (2009)
E042	alpha-bisabolol	515-69-5	3.69	222.372	Costa et al. (2004)
E043	alpha-cadinol	481-34-5	3.49	222.372	Costa et al. (2004)
E044	t-muurolol	19912-62-0	3.49	222.372	Costa et al. (2004)
E045	cis-isolongifolone	23787-90-8	4.15	220.356	Dias & Moraes (2014)
E046	delta-cadinene	483-76-1	4.14	204.357	Santos et al. (2006)
E047	tectoquinone	84-54-8	1.88	222.243	Cheng et al. (2008)
E048	guaiol	489-86-1	3.27	222.372	Dias & Moraes (2014)
E049	citronellic acid	502-47-6	2.62	170.252	Dias & Moraes (2014)
E050	myrtenol	515-00-4	1.84	152.237	Dias & Moraes (2014)
E051	16-kaurene	562-28-7	5.97	272.476	Cheng et al. (2009)
E052	elemol	639-99-6	3.84	222.372	Cheng et al. (2009)
E053	cedrol	77-53-2	3.57	222.372	Cheng et al. (2013)
E054	epi-zonarene	41702-63-0	4.14	204.357	Ali et al. (2013)
E055	beta-guaiene	88-84-6	3.98	204.357	Lima et al. (2013)
E056	ascaridole	512-85-6	1.99	168.236	Dias & Moraes (2014)
E057	spathulenol	6750-60-3	3.01	220.356	Lima et al. (2013)
E058	p-anisaldehyde	123-11-5	-0.45	136.150	Santos et al. (2007)
E059	m-eugenol	501-19-9	0.23	164.204	Dias & Moraes (2014)
E060	germacrene D	37839-63-7	4.69	204.357	Govindarajan (2010)
E061	benzyl benzoate	120-51-4	1.99	212.248	Jantan et al. (2005)
E062	methyl-cinnamate	103-26-4	1.57	162.188	Jantan et al. (2005)
E063	piperitone oxide	5286-38-4	1.85	168.236	Kulkarni et al. (2013)
E064	fenchone	1195-79-5	3.39	152.237	Kulkarni et al. (2013)
E065	cinnamaldehyde	104-55-2	1.04	132.162	Dias & Moraes (2014)
E066	alpha-phellandrene	2243-33-6	3.12	136.238	Jantan et al. (2005)

Table A1. (Continued).

Label	Name	CAS Number	log k_{ow}	Molecular Weight	Reference
E067	cinnamyl acetate	103-54-8	1.41	176.215	Dias & Moraes (2014)
E068	beta-phellandrene	555-10-2	3.17	136.238	Lucia et al. (2008)
E069	linalool	78-70-6	2.55	154.253	Dias & Moraes (2014)
E070	caryophyllene epoxide	17627-43-9	3.29	220.356	Magalhães et al. (2010)
E071	alpha-eudesmol	473-16-5	3.56	222.372	Lucia et al. (2012)
E072	p-menthane-3,8-diol	42822-86-6	1.48	172.268	Dias & Moraes (2014)
E073	citronellal	106-23-0	2.36	154.253	Waliwitiya et al. (2009)
E074	myristicin	607-91-0	-0.61	192.214	Dias & Moraes (2014)
E075	dillapiole	484-31-1	-1.58	222.240	Dias & Moraes (2014)
E076	alpha-copaene	3856-25-5	4.23	204.357	Magalhães et al. (2010)
E077	asaricin	18607-93-7	2.46	206.241	Dias & Moraes (2014)
E078	1-butyl-3,4-methylenedioxybenzene	NA	3.55	178.231	Nascimento et al. (2013)
E079	isoelemicin	487-12-7	-0.69	208.257	Costa et al. (2010)
E080	Z-asarone	5273-86-9	-0.69	208.257	Dias & Moraes (2014)
E081	patchouli alcohol	5986-55-0	3.85	222.372	Dias & Moraes (2014)
E082	alpha-asarone	494-40-6	-0.69	208.257	Dias & Moraes (2014)
E083	geijerene	6902-73-4	3.79	162.276	Dias & Moraes (2014)
E084	sabinene	3387-41-5	2.95	136.238	Dias & Moraes (2014)
E085	viridiflorol	552-02-3	3.43	222.372	Dias & Moraes (2014)
E086	bicyclogermacrene	24703-35-3	4.43	204.357	Costa et al. (2010)
E088	curcumene	644-30-4	4.17	202.341	Dias & Moraes (2014)
E089	ar-turmerone	532-65-0	3.57	216.324	Dias & Moraes (2014)
E090	zingiberene	495-60-3	4.64	204.357	Dias & Moraes (2014)
E091	beta-turmerone	82508-14-3	4.09	218.340	Dias & Moraes (2014)
E092	dodecanal	112-54-9	3.84	184.323	Dias & Moraes (2014)
E093	1-dodecanol	112-53-8	4.31	186.339	Dias & Moraes (2014)
E094	(E)-beta-ocimene	3779-61-1	3.28	136.238	Dias & Moraes (2014)
E095	myrcene epoxide	29414-55-9	2.07	152.237	Ruiz et al. (2011)
E096	dihydrotagetone	1879-00-1	3.15	154.253	Ruiz et al. (2011)
E097	t-cadinol	5937-11-1	3.49	222.372	Cheng et al. (2004)
E098	alpha-santalene	512-61-8	4.37	204.357	Magalhães et al. (2010)

Table A1. (Continued).

Label	Name	CAS Number	log k_{ow}	Molecular Weight	Reference
E099	neral	106-26-3	2.35	152.237	Ali et al. (2013)
E100	geranial	141-27-5	2.35	152.237	Dias & Moraes (2014)
E101	aromadendrene	25246-27-9	4.28	204.357	Morais et al. (2007)
E102	beta-selinene	17066-67-0	4.53	204.357	Morais et al. (2007)
E104	valencene	4630-07-3	4.48	204.357	Costa et al. (2010)
E105	fenchene	471-84-1	2.95	136.238	Perumalsamy et al. (2009)
E106	geranyl formate	105-86-2	2.45	182.263	Ali et al. (2013)
E107	(E),(E)-farnesol	4602-84-0	4.01	222.372	Dias & Moraes (2014)
E108	pregeijerene	NA	3.57	162.276	Dias & Moraes (2014)
E109	3,5- dimethoxytoluene	4179-19-5	-0.23	152.193	Perumalsamy et al. (2009)
E110	3,4,5- trimethoxytoluene	6443-69-2	-1.21	182.219	Perumalsamy et al. (2009)
E111	verbenone	80-57-9	2.54	150.221	Perumalsamy et al. (2009)
E112	para-methoxycinnamic acid	830-09-1	0.33	178.187	Dias & Moraes (2014)
E113	2,2-dimethyl-6-vinylchroman-4-one	79694-76-1	2.26	202.253	Albuquerque et al. (2004)
E114	2-senecioyl-4-vinylphenol	NA	2.99	202.253	Albuquerque et al. (2004)
E115	trans-ethyl cinnamate	103-36-6	1.91	176.215	Dias & Moraes (2014)
E116	hexyl butyrate	2639-63-6	3.1	172.268	Tabanca et al. (2012)
E117	benzyl salicylate	118-58-1	0.91	228.247	Jantan et al. (2005)
E118	ethyl-p-methoxycinnamate	1929-30-2	0.93	206.241	Dias & Moraes (2014)
E119	alpha-cedrene	469-61-4	4.37	204.357	Cheng et al. (2013)
E120	beta-cedrene	546-28-1	4.42	204.357	Cheng et al. (2013)
E121	octyl acetate	112-14-1	2.87	172.268	Tabanca et al. (2012)
E122	eucarvone	503-93-5	2.66	150.221	Perumalsamy et al. (2009)
E123	bornyl acetate	76-49-3	2.66	196.290	Waliwitiya et al. (2009)
E124	emodic acid	478-45-5	-2.3	300.222	Dias & Moraes (2014)
E126	guineensine	55038-30-7	3.73	383.532	Park et al. (2002)
E127	pipercide	54794-74-0	2.9	355.478	Park et al. (2002)
E128	retrofractamide A	94079-67-1	2.06	327.424	Park et al. (2002)
E129	(Z,Z)-matricaria ester	928-36-9	2.23	174.199	Cantrell et al. (2010)
E130	(E)-cinnamic acid	140-10-3	1.31	148.161	Cantrell et al. (2010)
E131	locustol	2785-88-8	0.08	152.193	Cantrell et al. (2010)

Table A1. (Continued).

Label	Name	CAS Number	log k_{ow}	Molecular Weight	Reference
E132	coumestrol	479-13-0	-3.14	268.224	Cantrell et al. (2010)
E133	parthenin	508-59-8	2.16	262.305	Cantrell et al. (2010)
E134	(+)-camphene	79-92-5	2.95	136.238	Santos et al. (2010)
E135	betulin	473-98-3	7.25	442.728	Cantrell et al. (2010)
E136	quercetin	117-39-5	-4.54	302.238	Cantrell et al. (2010)
E137	parthenolide	20554-84-1	2.07	248.322	Cantrell et al. (2010)
E138	rutin	153-18-4	-7.42	610.521	Cantrell et al. (2010)
E139	enhydrin	33880-85-2	0.24	464.467	Cantrell et al. (2010)
E140	ferulic acid	537-98-4	-0.75	194.186	Cantrell et al. (2010)
E141	(24R)-24,25-epoxycycloartan-3-one	NA	7.58	440.712	Cantrell et al. (2010)
E142	alantolactone	546-43-0	2.93	232.323	Cantrell et al. (2010)
E143	isoalantolactone	470-17-7	2.98	232.323	Cantrell et al. (2010)
E144	ergosterol endoperoxide	2061-64-5	6.37	428.657	Cantrell et al. (2010)
E145	pulegone	89-82-7	2.68	152.237	Waliwitiya et al. (2009)
E146	apiole	523-80-8	-1.58	222.240	Costa et al. (2010)
E148	beta-pinene	127-91-3	2.95	136.238	Lucia et al. (2007)
E149	beta-eudesmol	473-15-4	3.61	222.372	Cheng et al. (2009)
E150	cis-carveol	1197-06-4	1.92	152.237	Govindarajan et al. (2012)
E151	Z,E -nepetalactone	21651-62-7	1.97	166.220	Dias & Moraes (2014)
E152	E,Z- nepetalactone	17257-15-7	1.97	166.220	Dias & Moraes (2014)
E153	para-benzoquinone	106-51-4	1.61	108.096	Sousa et al. (2010)
E154	2-methyl parabenzoquinone	553-97-9	1.78	122.123	Sousa et al. (2010)
E155	2-isopropyl parabenzoquinone	15232-10-7	2.53	150.177	Sousa et al. (2010)

APPENDIX B: PREDICTED VS. OBSERVED GRAPHS OF THE SELECTED MODELS

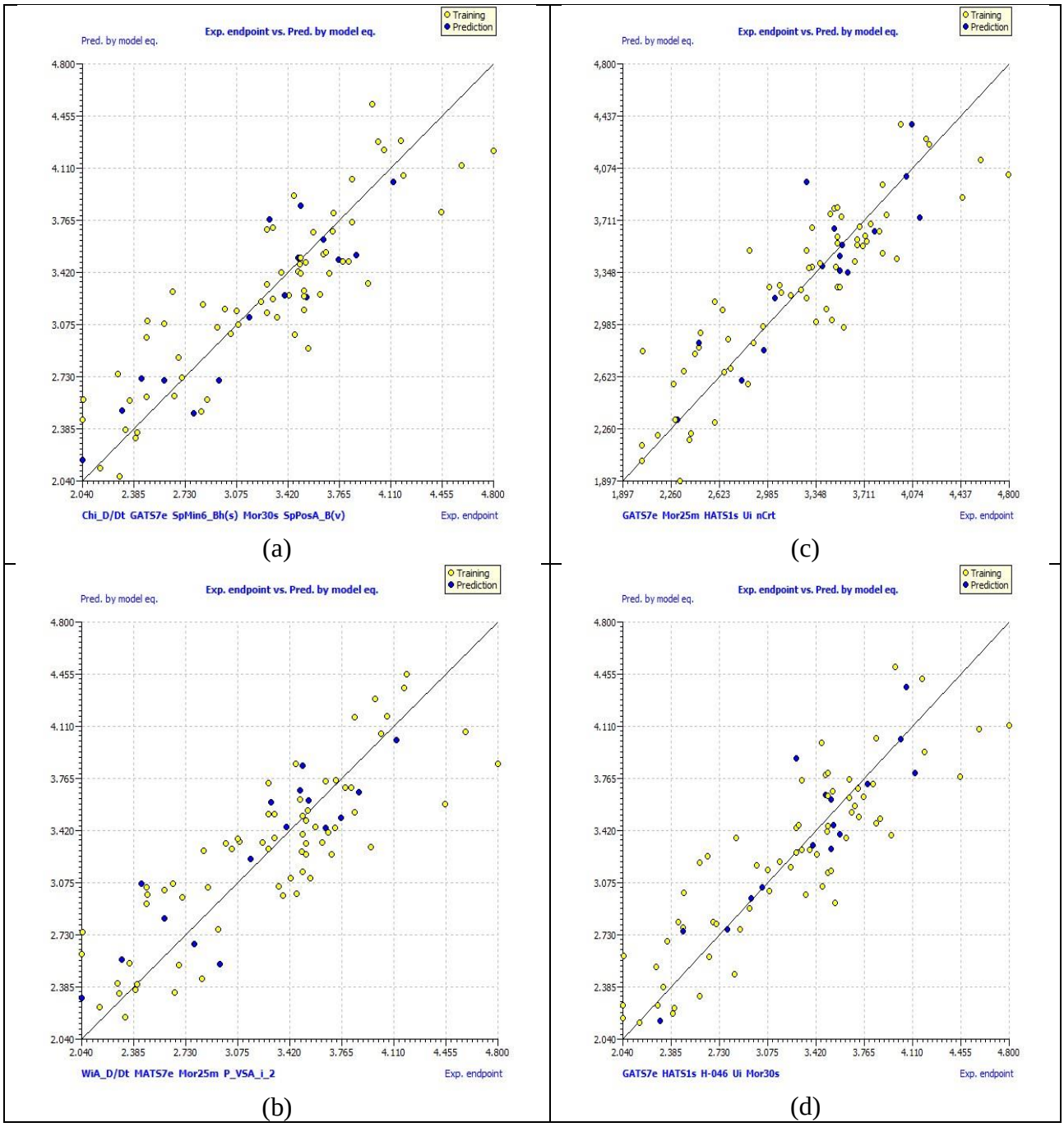

Figure B1. Predicted vs observed endpoints graphs of D1M1, D1M3, D2M1 and D2M2, respectively.

APPENDIX C. CHEMICALS USED IN MODELLING

Table C1: Chemicals used in modelling, training test set status, experimental and predicted toxicity values (Eq. 4.1), hat and descriptor values

Name	Status	Exp. pLC_{50} (M)	Pred. pLC_{50} (M) by Eq. 4.1	HAT i/i (*h**=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
carvacryl glycolic acid	Training	3.09	3.02	0.200	1.393	0.883	10	2.322	1.860
1,8-cineole	Training	2.04	2.18	0.186	0.000	0.389	4	0.000	0.977
1,4-cineole	Prediction	2.31	2.16	0.186	0.000	0.374	6	0.000	1.490
carvacrol	Training	3.47	3.05	0.048	0.160	0.680	10	2.000	0.174
carvacryl benzoate	Training	3.66	3.76	0.123	0.675	0.503	10	3.000	0.896
carvacryl acetate	Training	3.32	3.29	0.054	0.696	0.595	10	2.322	1.197
carvacryl chloroacetate	Training	3.64	3.37	0.044	1.060	0.621	10	2.322	1.144
2-hydroxy-3-methyl-6,-(1-methylethyl)-benzaldehyde	Training	3.43	3.26	0.165	0.097	0.769	10	2.322	-0.985
thymyl ethyl ether	Training	3.16	3.21	0.052	0.181	0.498	10	2.000	0.627
thymoxyacetic acid	Training	2.65	3.25	0.071	1.006	0.793	10	2.322	0.499
carvacryl propionate	Prediction	3.49	3.66	0.052	0.638	0.509	13	2.322	-0.003
carvacryl trichloroacetate	Prediction	3.59	3.40	0.077	1.537	0.765	10	2.322	0.746
thymyl acetate	Training	3.32	3.75	0.042	1.648	0.570	10	2.322	0.299
thymyl chloroacetate	Training	3.66	3.63	0.054	1.696	0.666	10	2.322	0.351
thymyl trichloroacetate	Training	3.85	3.47	0.099	1.830	0.772	10	2.322	0.783
thymyl propionate	Training	3.49	3.78	0.033	1.479	0.521	13	2.322	0.614
thymyl benzoate	Training	3.46	4.00	0.069	1.778	0.613	9	3.000	0.412
2-hydroxy-6-methyl-3-(1-methylethyl)-benzaldehyde	Training	3.72	3.69	0.166	1.722	0.771	10	2.322	-0.75
5-norbornene-2-ol	Training	2.16	2.15	0.082	0.000	0.721	3	1.000	1.209
5-norbornene-2,2-dimethanol	Training	2.29	2.26	0.067	0.000	0.700	6	1.000	1.135
5-norbornene-2-endo-3-endodimethanol	Training	2.04	2.26	0.061	0.000	0.715	4	1.000	0.735
5-norbornene-2-exo-3-exo-dimethanol	Training	2.33	2.38	0.083	0.000	0.732	4	1.000	-0.126
eugenyl acetate	Training	3.28	3.27	0.065	0.984	0.659	2	2.585	0.835

Table C1. (Continued).

Name	Status	Exp. pLC_{50} (M)	Pred. pLC_{50} (M) by Eq. 4.1	HAT i/i (*h**=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
2-(2-methoxy-4-(2-propen-1-yl))-phenoxy acetic acid	Prediction	3.04	3.04	0.123	1.123	0.777	2	2.585	1.582
borneol	Training	2.40	2.21	0.114	0.000	0.467	14	0.000	1.676
catechol	Training	2.66	2.58	0.055	0.000	0.767	0	2.000	0.532
alpha-terpinene	Training	3.76	3.64	0.046	1.625	0.446	14	1.585	0.443
terpineol	Training	3.68	3.54	0.272	2.854	0.457	9	1.000	0.782
1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	Prediction	3.40	3.32	0.086	0.687	0.514	2	2.322	0.416
eugenol	Training	3.35	3.00	0.058	1.041	0.817	2	2.322	0.684
phenol	Training	2.69	2.81	0.066	0.000	0.632	0	2.000	0.180
g-terpinene	Training	3.54	3.68	0.054	1.625	0.441	14	1.585	0.250
guaiacol	Training	2.84	2.47	0.070	0.000	0.834	0	2.000	0.737
1-benzoate-2-methoxy-4-(3-hydroxypropyl)-phenol	Training	3.28	3.44	0.251	1.131	0.563	2	3.000	1.964
4-hydroxy-3-methoxy-benzenepropanol	Training	2.05	2.59	0.143	1.03	0.941	2	2.000	1.272
isoborneol	Training	2.41	2.24	0.107	0.000	0.46	14	0.000	1.524
isopulegol	Training	2.71	2.80	0.037	0.217	0.517	11	1.000	0.422
thymol	Training	2.59	3.21	0.040	0.224	0.581	10	2.000	0.092
menthone	Training	2.48	3.01	0.046	0.180	0.392	15	1.000	0.644
nonan-2-one	Training	2.85	3.37	0.075	1.217	0.348	13	1.000	0.329
undecan-2-one	Training	3.51	3.45	0.062	0.986	0.29	17	1.000	0.487
1,2-dimethoxy-4-(2-propen-1-yl)-benzene	Training	3.24	3.18	0.046	0.795	0.636	2	2.322	0.547
neo-isopulegol	Training	2.44	2.81	0.037	0.217	0.502	11	1.000	0.462
1,2-carvone oxide	Training	2.88	2.76	0.036	0.162	0.649	4	1.585	0.102
limonene oxide, cis	Training	2.47	2.77	0.064	0.298	0.496	6	1.000	0.133
p-cymene	Training	3.51	3.65	0.040	1.577	0.512	10	2.000	0.355
eugenyl propionate	Prediction	3.55	3.45	0.086	1.18	0.573	5	2.585	1.197
R-carvone	Training	3.00	3.19	0.037	0.28	0.593	9	2.000	0.052
S-carvone	Training	3.08	3.16	0.035	0.28	0.61	9	2.000	0.105
R-limonene	Prediction	3.79	3.73	0.085	2.031	0.466	13	1.585	0.321

Table C1. (Continued).

Name	Status	Exp. pLC50 (M)	Pred. pLC50 (M) by Eq. 4.1	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
s-limonene	Training	3.83	3.73	0.084	2.031	0.466	13	1.585	0.338
resorcinol	Training	2.28	2.52	0.069	0.000	0.874	0	2.000	0.118
salicyl aldehyde	Training	2.95	2.90	0.066	0.000	0.741	0	2.322	-0.292
vanillin	Prediction	2.47	2.75	0.202	1.454	1.052	0	2.322	0.796
2,6-dimethyl-p-benzoquinone	Training	3.51	3.14	0.173	0.000	0.799	6	2.322	-1.269
2,5-dimethyl-p-benzoquinone	Training	3.38	3.29	0.149	0.000	0.677	6	2.322	-1.245
thymoquinone	Prediction	3.53	3.30	0.086	0.122	0.668	10	2.322	-0.416
pipilyasine	Training	3.99	4.50	0.113	2.156	0.252	24	2.000	0.291
pipzubedine	Training	4.18	4.42	0.131	1.873	0.248	32	2.000	1.499
pipyaqubine	Prediction	4.03	4.02	0.122	0.816	0.242	27	2.000	1.340
pellitorine	Prediction	4.07	4.37	0.126	2.577	0.299	18	2.000	0.589
pipericine	Prediction	4.13	3.80	0.205	0.741	0.233	39	1.000	1.640
piperine	Training	4.45	3.78	0.203	0.727	0.388	2	2.807	0.037
(-)-camphene	Prediction	2.79	2.77	0.048	0.000	0.461	14	1.000	1.113
3-carene	Prediction	2.96	2.97	0.055	0.000	0.407	15	1.000	0.463
camphor	Training	2.36	2.68	0.055	0.000	0.493	14	1.000	1.353
menthol	Training	2.59	2.32	0.102	0.191	0.498	15	0.000	1.240
tetradecanoic acid	Training	3.96	3.39	0.065	0.847	0.315	25	1.000	1.613
2,4-di-t-butylphenol	Training	4.80	4.11	0.086	2.338	0.438	18	2.000	0.657
linoleic acid	Training	4.59	4.09	0.099	0.882	0.254	25	2.000	0.646
nerolidol	Training	4.20	3.94	0.062	1.814	0.362	15	2.000	0.911
palmitic acid	Training	3.88	3.50	0.093	0.796	0.275	29	1.000	1.753
methyl linolelaidate	Training	3.85	4.03	0.093	0.855	0.273	25	2.000	0.832
caryophyllene	Prediction	3.53	3.62	0.048	1.128	0.35	21	1.585	1.432
geranic acid	Training	3.53	3.15	0.078	1.021	0.753	13	2.000	0.954
terpinen-4-ol	Training	3.56	2.94	0.057	0.21	0.452	11	1.000	0.075
ethyl palmitate	Training	3.73	3.51	0.093	0.763	0.264	29	1.000	1.719
humulene	Prediction	3.28	3.90	0.049	1.146	0.347	20	2.000	0.820
behenic acid	Training	3.51	3.80	0.249	0.725	0.19	41	1.000	2.246
n-hexadecane	Training	3.30	3.45	0.154	0.955	0.146	34	0.000	1.286

Table C1. (Continued).

Name	Status	Exp. pLC50 (M)	Pred. pLC50 (M) by Eq. 4.1	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
trans-anethole	Training	3.70	3.58	0.083	1.355	0.502	3	2.322	0.292
estragole	Training	3.50	3.41	0.061	1.549	0.656	2	2.322	0.303

APPENDIX D. EXTERNAL SET CHEMICALS PREDICTED BY MODELS

Table D1: Chemicals with their labels, their predicted pLC_{50} values from Eq. 4.1, hat values and descriptor values

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E001	g-elemene	3.86	0.040	1.372	0.455	19	2.000	0.497
E002	1-(cyclohexylacetyl)-2-methyl-piperidine	3.25	0.076	0.691	0.293	13	1.000	0.602
E003	(2R)-1-decanoyl-2-methyl-piperidine	3.36	0.042	0.812	0.331	19	1.000	0.750
E004	1-dodecanoyl-2-methyl-piperidine	3.47	0.052	0.781	0.280	23	1.000	1.032
E005	(2R)-1-heptanoyl-2-methyl-piperidine	3.20	0.047	0.898	0.390	13	1.000	0.530
E006	1-(3-cyclohexylpropanoyl)-2-methyl-piperidine	3.56	0.092	1.460	0.294	15	1.000	0.382
E007	1-[(4-methylcyclohexyl)-carbonyl]-2-methyl-piperidine	3.30	0.069	0.788	0.298	14	1.000	0.542
E008	(3S)-1-(1-methylcyclohexyl)-carbonyl-3-methyl-piperidine	3.33	0.058	0.602	0.275	18	1.000	0.796
E009	(3S)-1-(3-cyclohexylpropanoyl)-3-methyl-piperidine	3.61	0.079	1.470	0.324	18	1.000	0.314
E010	(3S)-1-heptanoyl-3-methyl-piperidine	3.45	0.087	0.898	0.340	16	1.000	-0.130
E011	(3S)-1-(cyclohexylcarbonyl)-3-methyl-piperidine	3.30	0.054	0.665	0.315	16	1.000	0.530
E012	1-decanoyl-4-methyl-piperidine	3.71	0.083	1.407	0.289	21	1.000	0.287
E013	1-(4-cyclohexylbutanoyl)-4-methyl-piperidine	3.85	0.132	2.184	0.283	19	1.000	0.563
E014	1-(cyclohexylcarbonyl)-4-methyl-piperidine	3.33	0.078	0.713	0.305	14	1.000	0.195
E015	1-(3-cyclohexylpropanoyl)-4-methyl-piperidine	3.77	0.135	2.195	0.302	17	1.000	0.627
E016	1-dodecanoyl-4-methyl-piperidine	3.81	0.095	1.324	0.270	25	1.000	0.283
E017	1-(cyclohexylcarbonyl)-2-ethyl-piperidine	3.13	0.076	0.166	0.298	15	1.000	0.664
E018	1-(3-cyclohexylpropanoyl)-2-ethyl-piperidine	3.40	0.051	1.303	0.332	18	1.000	1.190
E019	1-propionyl-2-ethyl-piperidine	2.89	0.087	0.067	0.406	8	1.000	0.054
E020	1-(3-cyclopentylpropanoyl)-2-ethyl-piperidine	3.20	0.073	1.172	0.351	16	1.000	1.729
E021	1-nonanoyl-2-ethyl-piperidine	3.32	0.038	0.764	0.350	20	1.000	0.901
E022	1-octanoyl-3-benzyl-piperidine	4.03	0.067	1.069	0.369	17	2.322	0.225
E023	1-undec-10-enoyl-4-benzyl-piperidine	4.38	0.138	1.278	0.343	17	2.585	-0.579
E024	1-cyclohexylacetyl-4-benzyl-piperidine	3.87	0.077	1.164	0.358	14	2.322	0.985
E025	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	4.20	0.071	1.816	0.356	16	2.322	0.423

Table D1. (Continued).

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E026	2-methyl-1-undec-10-enoyl-piperidine	3.58	0.051	0.809	0.314	16	1.585	0.704
E027	2-ethyl-1-undec-10-enoyl-piperidine	3.61	0.049	0.751	0.311	19	1.585	0.886
E028	2-benzyl-1-undec-10-enoyl-piperidine	4.10	0.106	1.193	0.323	16	2.585	0.925
E029	3-methyl-1-undec-10-enoyl-piperidine	3.63	0.044	0.809	0.389	19	1.585	0.295
E030	3-ethyl-1-undec-10-enoyl-piperidine	3.80	0.040	1.279	0.338	21	1.585	0.764
E031	3-benzyl-1-undec-10-enoyl-piperidine	4.19	0.096	1.038	0.352	18	2.585	0.202
E032	4-methyl-1-undec-10-enoyl-piperidine	3.97	0.088	1.395	0.272	18	1.585	0.006
E033	4-ethyl-1-undec-10-enoyl-piperidin	3.74	0.045	1.372	0.326	20	1.585	1.200
E035	alpha-pinene	2.89	0.047	0.000	0.432	15	1.000	0.770
E037	safrole	3.42	0.068	1.509	0.712	2	2.322	-0.168
E038	piperitone	3.20	0.040	0.238	0.511	14	1.585	0.202
E040	trans-ocimenone	3.40	0.062	0.237	0.553	9	2.322	-0.087
E041	terpinolene	3.83	0.076	2.031	0.394	15	1.585	0.568
E042	alpha-bisabolol	4.05	0.133	2.623	0.361	17	1.585	0.872
E043	alpha-cadinol	3.66	0.119	2.161	0.391	18	1.000	0.724
E044	t-muurolol	3.70	0.130	2.161	0.393	18	1.000	0.473
E045	cis-isolongifolone	3.05	0.076	0.000	0.351	21	1.000	1.314
E046	delta-cadinene	3.84	0.051	1.235	0.358	23	1.585	0.597
E047	tectoquinone	3.99	0.158	0.910	0.509	3	3.170	-0.686
E048	guaiol	3.70	0.120	2.155	0.362	18	1.000	0.665
E049	citronellic acid	3.06	0.062	0.855	0.698	14	1.585	0.660
E050	myrtenol	2.53	0.054	0.000	0.602	12	1.000	1.198
E051	16-kaurene	3.61	0.097	1.090	0.268	30	1.000	1.821
E052	elemol	3.81	0.126	2.485	0.45	13	1.585	0.757
E053	cedrol	3.17	0.215	2.079	0.317	19	0.000	1.479
E054	epi-zonarene	3.75	0.045	1.103	0.365	23	1.585	0.815
E055	beta-guaiene	3.85	0.054	1.118	0.323	24	1.585	0.725
E056	ascaridole	2.75	0.069	0.138	0.471	6	1.000	0.219
E057	spathulenol	2.95	0.050	0.048	0.395	15	1.000	0.796
E058	p-anisaldehyde	2.99	0.048	0.857	0.809	0	2.322	0.182

Table D1. (Continued).

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E059	m-eugenol	3.11	0.104	1.636	0.844	2	2.322	0.854
E060	germacrene D	3.92	0.044	1.441	0.365	19	2.000	0.966
E061	benzyl benzoate	3.42	0.107	0.736	0.637	0	3.000	0.544
E062	methyl-cinnamate	3.48	0.140	1.975	0.653	0	2.585	1.122
E063	piperitone oxide	2.77	0.049	0.138	0.459	9	1.000	0.661
E064	fenchone	2.83	0.055	0.000	0.435	16	1.000	1.236
E065	cinnamaldehyde	3.64	0.419	3.280	0.985	0	2.585	-0.076
E066	alpha-phellandrene	3.60	0.045	1.625	0.437	13	1.585	0.656
E067	cinnamyl acetate	3.35	0.065	1.216	0.671	0	2.585	0.458
E068	beta-phellandrene	3.53	0.049	1.693	0.480	12	1.585	0.677
E069	linalool	3.63	0.192	2.667	0.586	8	1.585	0.408
E070	caryophyllene epoxide	3.56	0.117	1.995	0.345	14	1.000	0.791
E071	alpha-eudesmol	4.01	0.272	2.911	0.348	17	1.000	0.086
E072	p-menthane-3,8-diol	2.57	0.191	1.205	0.494	8	0.000	0.543
E073	citronellal	3.16	0.055	0.915	0.667	14	1.585	0.418
E074	myristicin	3.24	0.052	0.852	0.725	2	2.322	-0.390
E075	dillapiole	3.32	0.056	0.691	0.619	2	2.322	-0.312
E076	alpha-copaene	3.47	0.049	1.116	0.360	23	1.000	1.002
E077	asaricin	3.36	0.057	1.294	0.616	2	2.322	0.529
E078	1-butyl-3,4-methylenedioxybenzene	3.54	0.043	1.288	0.576	9	2.000	-0.116
E079	isoelemicin	3.18	0.077	0.785	0.583	3	2.322	1.059
E080	Z-asarone	3.38	0.044	1.214	0.621	3	2.322	0.369
E081	patchouli alcohol	2.61	0.129	0.166	0.311	23	0.000	2.100
E082	alpha-asarone	3.55	0.079	1.214	0.501	3	2.322	0.279
E083	geijerene	3.18	0.058	0.000	0.521	11	2.000	0.469
E084	sabinene	2.82	0.043	0.000	0.465	14	1.000	0.777
E085	viridiflorol	2.96	0.132	1.323	0.336	19	0.000	1.243
E086	bicyclogermacrene	3.71	0.042	1.018	0.345	22	1.585	0.926
E088	curcumene	3.96	0.049	1.245	0.403	17	2.322	0.694
E089	ar-turmerone	3.90	0.064	1.033	0.436	13	2.585	0.532

Table D1. (Continued).

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E090	zingiberene	3.88	0.040	1.286	0.403	20	2.000	0.789
E091	beta-turmerone	3.84	0.044	1.326	0.445	15	2.322	0.910
E092	dodecanal	3.43	0.049	0.925	0.364	21	1.000	0.575
E093	1-dodecanol	3.03	0.111	0.891	0.249	21	0.000	1.029
E094	(E)-beta-ocimene	3.58	0.051	1.719	0.614	11	2.000	0.396
E095	myrcene epoxide	3.09	0.072	1.235	0.570	2	1.585	0.359
E096	dihydrotagetone	3.09	0.041	0.186	0.553	11	1.585	0.001
E097	t-cadinol	3.69	0.125	2.161	0.390	18	1.000	0.561
E098	alpha-santalene	3.57	0.057	1.334	0.350	23	1.000	0.869
E099	neral-citral	3.59	0.036	1.106	0.559	13	2.000	0.010
E100	geranial	3.53	0.040	1.106	0.597	13	2.000	0.035
E101	aromadendrene	3.09	0.082	0.278	0.341	22	1.000	1.760
E102	beta-selinene	3.83	0.047	1.583	0.393	20	1.585	0.575
E104	valencene	3.74	0.054	0.947	0.379	21	1.585	0.250
E105	fenchene	2.79	0.045	0.000	0.461	14	1.000	0.980
E106	geranyl formate	3.58	0.039	1.537	0.597	13	2.000	0.509
E107	(E),(E)-farnesol	3.87	0.067	0.780	0.364	20	2.000	0.228
E108	pregeijerene	3.74	0.048	0.974	0.437	14	2.000	-0.060
E109	3,5- dimethoxytoluene	2.82	0.056	0.029	0.635	3	2.000	0.675
E110	3,4,5- trimethoxytoluene	2.78	0.088	0.027	0.602	3	2.000	1.108
E111	verbenone	2.66	0.080	0.000	0.694	12	1.585	1.322
E112	para-methoxycinnamic acid	3.10	0.204	2.015	0.974	0	2.585	1.036
E113	2,2-dimethyl-6-vinylchroman-4-one	3.42	0.070	0.997	0.618	0	2.585	0.014
E114	2-senecioyl-4-vinylphenol	3.97	0.137	1.821	0.704	6	2.807	-0.932
E115	trans-ethyl cinnamate	3.45	0.128	1.653	0.605	0	2.585	1.081
E116	hexyl butyrate	3.27	0.054	1.133	0.352	14	1.000	0.928
E117	benzyl salicylate	3.36	0.074	0.725	0.728	0	3.000	0.219
E118	ethyl-p-methoxycinnamate	3.38	0.158	1.455	0.577	0	2.585	1.383
E119	alpha-cedrene	3.31	0.059	0.596	0.305	23	1.000	1.437
E120	beta-cedrene	3.25	0.053	0.561	0.331	22	1.000	1.377

Table D1. (Continued).

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E121	octyl acetate	3.08	0.040	1.085	0.467	13	1.000	0.959
E122	eucarvone	3.16	0.060	0.000	0.516	9	2.000	0.377
E123	bornyl acetate	3.28	0.088	1.743	0.425	14	1.000	1.387
E124	emodic acid	3.88	0.153	1.254	0.754	0	3.322	-1.270
E126	guineensine	4.54	0.247	0.798	0.319	18	3.000	-0.95
E127	pipercide	4.30	0.166	0.844	0.379	14	3.000	-0.504
E128	retrofractamide A	4.16	0.143	0.899	0.425	10	3.000	-0.478
E129	(Z,Z)-matricaria ester	3.57	0.082	0.648	0.635	3	3.000	-0.032
E130	(E)-cinnamic acid	2.98	0.467	2.565	1.187	0	2.585	1.082
E131	locustol	3.11	0.095	1.665	0.792	5	2.000	0.862
E132	coumestrol	3.71	0.133	0.760	0.615	0	3.322	-0.098
E133	parthenin	3.49	0.044	1.462	0.658	10	2.322	0.906
E134	(+)-camphene	2.77	0.048	0.000	0.461	14	1.000	1.111
E135	betulin	3.68	0.313	1.023	0.260	41	1.000	2.985
E136	quercetin	2.77	0.405	0.823	1.313	0	3.170	-0.054
E137	parthenolide	3.36	0.051	1.483	0.585	9	2.000	1.250
E138	rutin	4.09	0.384	0.928	0.641	0	3.170	-2.678
E139	enhydrin	3.77	0.133	1.064	0.470	2	2.807	0.086
E140	ferulic acid	2.36	0.361	0.826	1.201	0	2.585	1.674
E141	(24R)-24,25-epoxycycloartan-3-one	3.78	0.171	0.763	0.205	37	1.000	1.784
E142	alantolactone	3.87	0.137	2.628	0.508	13	2.000	1.377
E143	isoalantolactone	3.57	0.277	2.594	0.540	12	2.000	2.711
E144	ergosterol endoperoxide	3.67	0.282	0.696	0.217	32	1.585	3.032
E145	pulegone	3.33	0.067	0.199	0.443	14	1.585	-0.161
E146	apiole	3.37	0.054	1.027	0.652	2	2.322	-0.313
E148	beta-pinene	2.86	0.046	0.000	0.479	14	1.000	0.462
E149	beta-eudesmol	3.91	0.249	2.911	0.366	16	1.000	0.446
E150	cis-carveol	2.87	0.026	0.205	0.615	9	1.585	0.550
E151	Z,E- nepetalactone	2.71	0.067	0.000	0.674	12	1.585	1.178

Table D1. (Continued).

Label	Name	pLC_{50} (Eq. 4.1)	HAT i/i (h*=0.272)	GATS7e	HATS1s	H-046	Ui	Mor30s
E152	E,Z- nepetalactone	2.80	0.053	0.000	0.610	12	1.585	1.123
E153	para-benzoquinone	2.89	0.079	0.000	0.790	0	2.322	-0.572
E154	2-methyl parabenzoquinone	2.91	0.130	0.000	0.875	3	2.322	-0.849
E155	2-isopropyl parabenzoquinone	2.87	0.134	0.000	0.897	7	2.322	-0.224

Table D2. Chemicals with their labels, their predicted pLC_{50} values from Eq. 4.3, hat values and descriptor values

Label	Name	pLC_{50}	Hat values (*h**=0.219)	GATS7e	HATS1s	H-046	Ui	Mor30s
E001	g-elemene	3.85	0.031	1.372	0.455	19	2.000	0.497
E002	1-(cyclohexylacetyl)-2-methyl-piperidine	3.24	0.057	0.691	0.293	13	1.000	0.602
E003	(2R)-1-decanoyl-2-methyl-piperidine	3.36	0.032	0.812	0.331	19	1.000	0.750
E004	1-dodecanoyl-2-methyl-piperidine	3.47	0.038	0.781	0.280	23	1.000	1.032
E005	(2R)-1-heptanoyl-2-methyl-piperidine	3.19	0.038	0.898	0.390	13	1.000	0.530
E006	1-(3-cyclohexylpropanoyl)-2-methyl-piperidine	3.54	0.074	1.460	0.294	15	1.000	0.382
E007	1-[(4-methylcyclohexyl) carbonyl]-2-methyl-piperidine	3.30	0.052	0.788	0.298	14	1.000	0.542
E008	(3S)-1-(1-methylcyclohexyl) carbonyl-3-methyl-piperidine	3.34	0.043	0.602	0.275	18	1.000	0.796
E009	(3S)-1-(3-cyclohexylpropanoyl)-3-methyl-piperidine	3.59	0.066	1.470	0.324	18	1.000	0.314
E010	(3S)-1-heptanoyl-3-methyl-piperidine	3.44	0.071	0.898	0.340	16	1.000	-0.130
E011	(3S)-1-(cyclohexylcarbonyl)-3-methyl-piperidine	3.30	0.042	0.665	0.315	16	1.000	0.530
E012	1-decanoyl-4-methyl-piperidine	3.70	0.068	1.407	0.289	21	1.000	0.287
E013	1-(4-cyclohexylbutanoyl)-4-methyl-piperidine	3.82	0.109	2.184	0.283	19	1.000	0.563
E014	1-(cyclohexylcarbonyl)-4-methyl-piperidine	3.33	0.061	0.713	0.305	14	1.000	0.195
E015	1-(3-cyclohexylpropanoyl)-4-methyl-piperidine	3.74	0.111	2.195	0.302	17	1.000	0.627
E016	1-dodecanoyl-4-methyl-piperidine	3.80	0.076	1.324	0.270	25	1.000	0.283
E017	1-(cyclohexylcarbonyl)-2-ethyl-piperidine	3.13	0.055	0.166	0.298	15	1.000	0.664
E018	1-(3-cyclohexylpropanoyl)-2-ethyl-piperidine	3.39	0.041	1.303	0.332	18	1.000	1.190
E019	1-propionyl-2-ethyl-piperidine	2.90	0.068	0.067	0.406	8	1.000	0.054
E020	1-(3-cyclopentylpropanoyl)-2-ethyl-piperidine	3.19	0.057	1.172	0.351	16	1.000	1.729
E021	1-nonanoyl-2-ethyl-piperidine	3.32	0.030	0.764	0.350	20	1.000	0.901
E022	1-octanoyl-3-benzyl-piperidine	4.01	0.050	1.069	0.369	17	2.322	0.225
E023	1-undec-10-enoyl-4-benzyl-piperidine	4.35	0.110	1.278	0.343	17	2.585	-0.579
E024	1-cyclohexylacetyl-4-benzyl-piperidine	3.85	0.057	1.164	0.358	14	2.322	0.985
E025	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	4.17	0.055	1.816	0.356	16	2.322	0.423
E026	2-methyl-1-undec-10-enoyl-piperidine	3.57	0.037	0.809	0.314	16	1.585	0.704
E027	2-ethyl-1-undec-10-enoyl-piperidine	3.61	0.035	0.751	0.311	19	1.585	0.886
E028	2-benzyl-1-undec-10-enoyl-piperidine	4.08	0.080	1.193	0.323	16	2.585	0.925
E029	3-methyl-1-undec-10-enoyl-piperidine	3.62	0.033	0.809	0.389	19	1.585	0.295

Table D2. (Continued).

Label	Name	pLC_{50}	Hat values (*h**=0.219)	GATS7e	HATS1s	H-046	Ui	Mor30s
E030	3-ethyl-1-undec-10-enoyl-piperidine	3.79	0.030	1.279	0.338	21	1.585	0.764
E031	3-benzyl-1-undec-10-enoyl-piperidine	4.17	0.072	1.038	0.352	18	2.585	0.202
E032	4-methyl-1-undec-10-enoyl-piperidine	3.95	0.070	1.395	0.272	18	1.585	0.006
E033	4-ethyl-1-undec-10-enoyl-piperidin	3.73	0.033	1.372	0.326	20	1.585	1.200
E035	alpha-pinene	2.91	0.036	0.000	0.432	15	1.000	0.770
E037	safrole	3.39	0.054	1.509	0.712	2	2.322	-0.168
E038	piperitone	3.21	0.032	0.238	0.511	14	1.585	0.202
E040	trans-ocimenone	3.40	0.050	0.237	0.553	9	2.322	-0.087
E041	terpinolene	3.80	0.062	2.031	0.394	15	1.585	0.568
E042	alpha-bisabolol	4.00	0.107	2.623	0.361	17	1.585	0.872
E043	alpha-cadinol	3.63	0.099	2.161	0.391	18	1.000	0.724
E044	t-muurolol	3.67	0.108	2.161	0.393	18	1.000	0.473
E045	cis-isolongifolone	3.07	0.055	0.000	0.351	21	1.000	1.314
E046	delta-cadinene	3.83	0.038	1.235	0.358	23	1.585	0.597
E047	tectoquinone	3.97	0.131	0.910	0.509	3	3.170	-0.686
E048	guaiol	3.67	0.099	2.155	0.362	18	1.000	0.665
E049	citronellic acid	3.07	0.049	0.855	0.698	14	1.585	0.660
E050	myrtenol	2.55	0.045	0.000	0.602	12	1.000	1.198
E051	16-kaurene	3.62	0.074	1.090	0.268	30	1.000	1.821
E052	elemol	3.77	0.100	2.485	0.450	13	1.585	0.757
E053	cedrol	3.15	0.179	2.079	0.317	19	0.000	1.479
E054	epi-zonarene	3.75	0.033	1.103	0.365	23	1.585	0.815
E055	beta-guaiene	3.85	0.039	1.118	0.323	24	1.585	0.725
E056	ascaridole	2.76	0.056	0.138	0.471	6	1.000	0.219
E057	spathulenol	2.96	0.038	0.048	0.395	15	1.000	0.796
E058	p-anisaldehyde	2.98	0.038	0.857	0.809	0	2.322	0.182
E059	m-eugenol	3.09	0.074	1.636	0.844	2	2.322	0.854
E060	germacrene D	3.90	0.033	1.441	0.365	19	2.000	0.966
E061	benzyl benzoate	3.41	0.086	0.736	0.637	0	3.000	0.544
E062	methyl-cinnamate	3.44	0.103	1.975	0.653	0	2.585	1.122

Table D2. (Continued).

Label	Name	pLC_{50}	Hat values (*h**=0,2195)	GATS7e	HATS1s	H-046	Ui	Mor30s
E063	piperitone oxide	2.77	0.039	0.138	0.459	9	1.000	0.661
E064	fenchone	2.85	0.042	0.000	0.435	16	1.000	1.236
E065	cinnamaldehyde	3.58	0.308	3.280	0.985	0	2.585	-0.076
E066	alpha-phellandrene	3.57	0.036	1.625	0.437	13	1.585	0.656
E067	cinnamyl acetate	3.32	0.050	1.216	0.671	0	2.585	0.458
E068	beta-phellandrene	3.51	0.039	1.693	0.48	12	1.585	0.677
E069	linalool	3.58	0.150	2.667	0.586	8	1.585	0.408
E070	caryophyllene epoxide	3.53	0.094	1.995	0.345	14	1.000	0.791
E071	alpha-eudesmol	3.96	0.224	2.911	0.348	17	1.000	0.086
E072	p-menthane-3,8-diol	2.57	0.157	1.205	0.494	8	0.000	0.543
E073	citronellal	3.16	0.044	0.915	0.667	14	1.585	0.418
E074	myristicin	3.23	0.045	0.852	0.725	2	2.322	-0.390
E075	dillapiole	3.30	0.048	0.691	0.619	2	2.322	-0.312
E076	alpha-copaene	3.47	0.039	1.116	0.36	23	1.000	1.002
E077	asaricin	3.33	0.043	1.294	0.616	2	2.322	0.529
E078	1-butyl-3,4-methylenedioxybenzene	3.52	0.036	1.288	0.576	9	2.000	-0.116
E079	isoelemicin	3.17	0.059	0.785	0.583	3	2.322	1.059
E080	Z-asarone	3.36	0.034	1.214	0.621	3	2.322	0.369
E081	patchouli alcohol	2.64	0.105	0.166	0.311	23	0.000	2.100
E082	alpha-asarone	3.52	0.062	1.214	0.501	3	2.322	0.279
E083	geijerene	3.19	0.045	0.000	0.521	11	2.000	0.469
E084	sabinene	2.84	0.034	0.000	0.465	14	1.000	0.777
E085	viridiflorol	2.96	0.112	1.323	0.336	19	0.000	1.243
E086	bicyclogermacrene	3.71	0.030	1.018	0.345	22	1.585	0.926
E088	curcumene	3.94	0.036	1.245	0.403	17	2.322	0.694
E089	ar-turmerone	3.88	0.049	1.033	0.436	13	2.585	0.532
E090	zingiberene	3.87	0.029	1.286	0.403	20	2.000	0.789
E091	beta-turmerone	3.82	0.033	1.326	0.445	15	2.322	0.91
E092	dodecanal	3.43	0.039	0.925	0.364	21	1.000	0.575
E093	1-dodecanol	3.04	0.092	0.891	0.249	21	0.000	1.029

Table D2. (Continued).

Label	Name	pLC_{50}	Hat values (*h**=0.219)	GATS7e	HATS1s	H-046	Ui	Mor30s
E094	(E)-beta-ocimene	3.56	0.040	1.719	0.614	11	2.000	0.396
E095	myrcene epoxide	3.07	0.056	1.235	0.57	2	1.585	0.359
E096	dihydrotagetone	3.10	0.035	0.186	0.553	11	1.585	0.001
E097	t-cadinol	3.66	0.104	2.161	0.39	18	1.000	0.561
E098	alpha-santalene	3.56	0.046	1.334	0.35	23	1.000	0.869
E099	neral	3.57	0.030	1.106	0.559	13	2.000	0.01
E100	geranial	3.52	0.033	1.106	0.597	13	2.000	0.035
E101	aromadendrene	3.11	0.062	0.278	0.341	22	1.000	1.76
E102	beta-selinene	3.81	0.038	1.583	0.393	20	1.585	0.575
E104	valencene	3.73	0.041	0.947	0.379	21	1.585	0.25
E105	fenchene	2.81	0.035	0.000	0.461	14	1.000	0.98
E106	geranyl formate!	3.56	0.030	1.537	0.597	13	2.000	0.509
E107	(E),(E)-farnesol	3.87	0.049	0.78	0.364	20	2.000	0.228
E108	pregeijerene	3.73	0.038	0.974	0.437	14	2.000	-0.06
E109	3,5- dimethoxytoluene	2.82	0.045	0.029	0.635	3	2.000	0.675
E110	3,4,5- trimethoxytoluene	2.79	0.069	0.027	0.602	3	2.000	1.108
E111	verbenone	2.68	0.065	0.000	0.694	12	1.585	1.322
E112	para-methoxycinnamic acid	3.07	0.145	2.015	0.974	0	2.585	1.036
E113	2,2-dimethyl-6-vinylchroman-4-one	3.40	0.056	0.997	0.618	0	2.585	0.014
E114	2-senecioyl-4-vinylphenol	3.93	0.113	1.821	0.704	6	2.807	-0.932
E115	trans-ethyl cinnamate	3.42	0.096	1.653	0.605	0	2.585	1.081
E116	hexyl butyrate	3.26	0.043	1.133	0.352	14	1.000	0.928
E117	benzyl salicylate	3.34	0.061	0.725	0.728	0	3.000	0.219
E118	ethyl-p-methoxycinnamate	3.35	0.119	1.455	0.577	0	2.585	1.383
E119	alpha-cedrene	3.32	0.043	0.596	0.305	23	1.000	1.437
E120	beta-cedrene	3.26	0.039	0.561	0.331	22	1.000	1.377
E121	octyl acetate	3.08	0.033	1.085	0.467	13	1.000	0.959
E122	eucarvone	3.16	0.046	0.000	0.516	9	2.000	0.377
E123	bornyl acetate	3.26	0.071	1.743	0.425	14	1.000	1.387
E124	emodic acid	3.85	0.131	1.254	0.754	0	3.322	-1.27

Table D2. (Continued).

Label	Name	pLC_{50}	Hat values (*h**=0.219)	GATS7e	HATS1s	H-046	Ui	Mor30s
E126	guineensine	4.52	0.192	0.798	0.319	18	3.000	-0.95
E127	pipercide	4.28	0.131	0.844	0.379	14	3.000	-0.504
E128	retrofractamide A	4.14	0.115	0.899	0.425	10	3.000	-0.478
E129	(Z,Z)-matricaria ester	3.56	0.068	0.648	0.635	3	3.000	-0.032
E130	(E)-cinnamic acid	2.95	0.335	2.565	1.187	0	2.585	1.082
E131	locustol	3.09	0.069	1.665	0.792	5	2.000	0.862
E132	coumestrol	3.69	0.109	0.760	0.615	0	3.322	-0.098
E133	parthenin	3.47	0.033	1.462	0.658	10	2.322	0.906
E134	(+)-camphene	2.79	0.037	0.000	0.461	14	1.000	1.111
E135	betulin	3.70	0.244	1.023	0.260	41	1.000	2.985
E136	quercetin	2.77	0.303	0.823	1.313	0	3.170	-0.054
E137	parthenolide	3.34	0.038	1.483	0.585	9	2.000	1.25
E138	rutin	4.06	0.328	0.928	0.641	0	3.170	-2.678
E139	enhydrin	3.74	0.105	1.064	0.470	2	2.807	0.086
E140	ferulic acid	2.36	0.271	0.826	1.201	0	2.585	1.674
E141	(24R)-24,25-epoxycycloartan-3-one	3.79	0.125	0.763	0.205	37	1.000	1.784
E142	alantolactone	3.82	0.105	2.628	0.508	13	2.000	1.377
E143	isoalantolactone	3.53	0.214	2.594	0.540	12	2.000	2.711
E144	ergosterol endoperoxide	3.69	0.218	0.696	0.217	32	1.585	3.032
E145	pulegone	3.34	0.052	0.199	0.443	14	1.585	-0.161
E146	apiole	3.35	0.046	1.027	0.652	2	2.322	-0.313
E148	beta-pinene	2.87	0.037	0.000	0.479	14	1.000	0.462
E149	beta-eudesmol	3.86	0.203	2.911	0.366	16	1.000	0.446
E150	cis-carveol	2.88	0.022	0.205	0.615	9	1.585	0.55
E151	Z,E- nepetalactone	2.73	0.054	0.000	0.674	12	1.585	1.178
E152	E,Z- nepetalactone	2.82	0.043	0.000	0.610	12	1.585	1.123
E153	para-benzoquinone	2.90	0.069	0.000	0.790	0	2.322	-0.572
E154	2-methyl parabenzoquinone	2.92	0.110	0.000	0.875	3	2.322	-0.849
E155	2-isopropyl parabenzoquinone	2.89	0.108	0.000	0.897	7	2.322	-0.224

APPENDIX E: PREDICTION OF AQUATIC TOXICITY MODELS

Table E1: Chemicals with their labels, their predicted pLC_{50} values from Eq. 4.9 (algae), hat values and descriptor values

Label	Chemical	pEC_{50}	Hat value (h^*= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
E001	g-elemene	6.15	0.050	0.355	0.718	1	0	0	0	20.560	5.035
E002	1-(cyclohexylacetyl)-2-methyl-piperidine	4.90	0.035	0.367	0.707	0	0	0	2	8.943	5.290
E003	(2R)-1-decanoyl-2-methyl-piperidine	6.86	0.105	0.441	1.052	0	0	0	2	19.431	5.285
E004	1-dodecanoyl-2-methyl-piperidine	7.03	0.093	0.388	1.019	0	0	0	2	23.901	5.305
E005	(2R)-1-heptanoyl-2-methyl-piperidine	5.10	0.051	0.426	0.777	0	0	0	2	7.436	5.285
E006	1-(3-cyclohexylpropanoyl)-2-methyl-piperidine	5.51	0.066	0.378	0.922	0	0	0	2	10.541	5.290
E007	1-[(4-methylcyclohexyl)carbonyl]-2-methyl-piperidine	5.21	0.044	0.381	0.758	0	0	0	3	8.943	5.280
E008	(3S)-1-(1-methylcyclohexyl)carbonyl-3-methyl-piperidine	5.31	0.046	0.351	0.786	0	0	0	3	10.541	5.285
E009	(3S)-1-(3-cyclohexylpropanoyl)-3-methyl-piperidine	5.53	0.056	0.374	0.853	0	0	0	3	10.541	5.265
E010	(3S)-1-heptanoyl-3-methyl-piperidine	5.23	0.050	0.435	0.744	0	0	0	3	7.436	5.270
E011	(3S)-1-(cyclohexylcarbonyl)-3-methyl-piperidine	5.25	0.049	0.368	0.801	0	0	0	3	8.943	5.275
E012	1-decanoyl-4-methyl-piperidine	6.29	0.064	0.361	0.903	0	0	0	2	19.431	5.290
E013	1-(4-cyclohexylbutanoyl)-4-methyl-piperidine	5.66	0.063	0.383	0.909	0	0	0	2	12.223	5.300
E014	1-(cyclohexylcarbonyl)-4-methyl-piperidine	5.06	0.043	0.389	0.743	0	0	0	3	7.436	5.280
E015	1-(3-cyclohexylpropanoyl)-4-methyl-piperidine	5.56	0.069	0.384	0.934	0	0	0	2	10.541	5.285
E016	1-dodecanoyl-4-methyl-piperidine	7.44	0.128	0.45	1.111	0	0	0	2	23.901	5.285
E017	1-(cyclohexylcarbonyl)-2-ethyl-piperidine	5.43	0.056	0.356	0.833	0	0	0	4	8.943	5.265
E018	1-(3-cyclohexylpropanoyl)-2-ethyl-piperidine	5.83	0.069	0.365	0.936	0	0	0	3	12.223	5.265
E019	1-propionyl-2-ethyl-piperidine	4.23	0.027	0.363	0.559	0	0	0	3	3.544	5.280
E020	1-(3-cyclopentylpropanoyl)-2-ethyl-piperidine	5.11	0.034	0.35	0.685	0	0	0	3	10.541	5.290
E021	1-nonanoyl-2-ethyl-piperidine	6.75	0.083	0.426	0.941	0	0	0	3	19.431	5.270
E022	1-octanoyl-3-benzyl-piperidine	7.54	0.085	0.389	0.861	0	0	0	3	25.545	4.655

Table E1. (Continued).

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
E023	1-undec-10-enoyl-4-benzyl-piperidine	7.52	0.117	0.332	1.114	1	0	0	2	22.211	4.740
E024	1-cyclohexylacetyl-4-benzyl-piperidine	6.48	0.069	0.380	0.863	0	0	0	2	17.158	4.730
E025	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	6.67	0.070	0.379	0.867	0	0	0	2	19.055	4.735
E026	2-methyl-1-undec-10-enoyl-piperidine	6.40	0.088	0.453	0.962	1	0	0	2	13.190	5.225
E027	2-ethyl-1-undec-10-enoyl-piperidine	6.96	0.118	0.444	1.091	1	0	0	3	14.972	5.195
E028	2-benzyl-1-undec-10-enoyl-piperidine	7.38	0.091	0.327	0.953	1	0	0	3	22.211	4.695
E029	3-methyl-1-undec-10-enoyl-piperidine	6.56	0.095	0.435	1.008	1	0	0	3	13.190	5.240
E030	3-ethyl-1-undec-10-enoyl-piperidine	6.88	0.108	0.441	1.054	1	0	0	3	14.972	5.195
E031	3-benzyl-1-undec-10-enoyl-piperidine	7.98	0.133	0.410	1.100	1	0	0	3	22.211	4.655
E032	4-methyl-1-undec-10-enoyl-piperidine	6.70	0.124	0.463	1.111	1	0	0	2	13.190	5.250
E033	4-ethyl-1-undec-10-enoyl-piperidin	6.76	0.109	0.435	1.082	1	0	0	2	14.972	5.200
E035	alpha-pinene	4.41	0.019	0.371	0.331	1	0	0	0	11.387	5.245
E037	safrole	4.33	0.010	0.49	0.063	1	0	0	0	4.512	4.325
E038	piperitone	4.45	0.023	0.398	0.478	1	0	0	0	5.062	4.865
E040	trans-ocimenone	4.89	0.062	0.455	0.255	3	0	0	0	6.027	4.730
E041	terpinolene	4.57	0.015	0.400	0.374	1	0	0	0	10.673	5.175
E042	alpha-bisabolol	5.48	0.055	0.359	0.685	2	0	0	0	13.061	5.205
E043	alpha-cadinol	5.36	0.046	0.341	0.792	1	0	0	0	13.849	5.335
E044	t-muurolol	5.41	0.050	0.339	0.824	1	0	0	0	13.849	5.335
E045	cis-isolongifolone	4.45	0.026	0.321	0.457	0	0	0	0	13.849	5.280
E046	delta-cadinene	6.18	0.054	0.351	0.755	1	0	0	0	21.446	5.190
E047	tectoquinone	4.61	0.007	0.454	0.053	0	0	0	0	9.912	4.160
E048	guaiol	5.25	0.047	0.338	0.798	0	0	0	0	13.849	5.175
E049	citronellic acid	4.27	0.014	0.391	0.330	1	0	0	0	6.621	4.905
E050	myrtenol	3.74	0.019	0.360	0.276	1	0	0	0	5.558	5.180
E051	16-kaurene	6.99	0.103	0.327	0.789	0	0	0	0	35.147	5.550
E052	elemol	4.99	0.034	0.345	0.662	1	0	0	0	13.061	5.415
E053	cedrol	3.95	0.042	0.332	0.580	0	0	0	0	14.95	6.430
E054	epi-zonarene	6.41	0.051	0.348	0.620	1	0	0	0	21.446	4.530
E055	beta-guaiene	6.08	0.053	0.349	0.783	0	0	0	0	21.446	5.075
E056	ascaridole	4.88	0.044	0.387	0.429	2	0	0	0	5.238	4.470
E057	spathulenol	4.64	0.027	0.335	0.592	0	0	0	0	13.849	5.440
E058	p-anisaldehyde	3.95	0.011	0.504	-0.025	1	0	0	0	2.219	4.380
E059	m-eugenol	4.30	0.010	0.470	0.085	1	0	0	0	4.453	4.330
E060	germacrene D	6.65	0.112	0.342	0.767	3	0	0	0	20.560	5.065
E061	benzyl benzoate	4.48	0.008	0.451	0.011	0	0	0	0	12.347	4.520

Table E1. (Continued).

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
E062	methyl-cinnamate	4.53	0.028	0.495	0.036	2	0	0	0	5.291	4.430
E063	piperitone oxide	3.76	0.024	0.391	0.457	0	0	0	0	2.181	5.005
E064	fenchone	3.59	0.015	0.357	0.290	0	0	0	0	5.558	5.085
E065	cinnamaldehyde	4.67	0.062	0.49	0.031	3	0	0	0	4.267	4.400
E066	alpha-phellandrene	5.47	0.071	0.404	0.382	3	0	0	0	10.673	4.655
E067	cinnamyl acetate	4.51	0.028	0.469	0.078	2	0	0	0	6.717	4.635
E068	beta-phellandrene	5.09	0.035	0.403	0.359	2	0	0	0	10.673	4.785
E069	linalool	4.66	0.034	0.417	0.440	2	0	0	0	6.980	5.140
E070	caryophyllene epoxide	4.96	0.038	0.339	0.744	0	0	0	0	13.849	5.415
E071	alpha-eudesmol	5.38	0.046	0.347	0.785	1	0	0	0	13.849	5.315
E072	p-menthane-3,8-diol	2.97	0.025	0.381	0.525	0	0	0	0	2.630	6.225
E073	citronellal	4.88	0.035	0.433	0.442	2	0	0	0	6.980	4.930
E074	myristicin	4.05	0.010	0.461	0.039	1	0	0	0	3.414	4.365
E075	dillapiole	3.97	0.014	0.436	0.037	1	0	0	0	2.525	4.240
E076	alpha-copaene	6.06	0.050	0.354	0.654	1	0	0	0	22.453	5.230
E077	asaricin	4.14	0.016	0.428	0.089	1	0	0	0	2.968	4.160
E078	1-butyl-3,4-methylenedioxybenzene	4.48	0.012	0.485	0.185	0	0	0	0	6.280	4.330
E079	isoelemicin	4.23	0.032	0.438	-0.022	2	0	0	0	4.473	4.315
E080	Z-asarone	4.24	0.039	0.396	-0.005	2	0	0	0	4.473	4.155
E081	patchouli alcohol	4.20	0.042	0.324	0.683	0	0	0	0	14.95	6.325
E082	alpha-asarone	4.48	0.036	0.421	0.001	2	0	0	0	4.473	3.965
E083	geijerene	5.35	0.082	0.374	0.479	3	0	0	0	14.008	5.345
E084	sabinene	4.19	0.012	0.395	0.355	0	0	0	0	11.387	5.385
E085	viridiflorol	4.35	0.044	0.326	0.760	0	0	0	0	14.950	6.335
E086	bicyclogermacrene	6.22	0.068	0.345	0.606	2	0	0	0	21.446	5.040
E088	curcumene	6.30	0.042	0.395	0.453	1	0	0	0	23.386	4.715
E089	ar-turmerone	5.45	0.021	0.419	0.395	1	0	0	0	13.935	4.580
E090	zingiberene	7.03	0.152	0.395	0.579	4	0	0	0	20.560	4.650
E091	beta-turmerone	5.69	0.074	0.406	0.420	3	0	0	0	11.744	4.610
E092	dodecanal	6.26	0.062	0.457	0.811	1	0	0	0	17.926	5.255
E093	1-dodecanol	5.27	0.070	0.469	0.851	0	0	0	0	19.175	6.560
E094	(E)-beta-ocimene	5.46	0.071	0.433	0.323	3	0	0	0	12.690	4.920
E095	myrcene epoxide	4.16	0.013	0.419	0.356	1	0	0	0	5.062	5.040
E096	dihydrotagetone	4.37	0.018	0.417	0.490	1	0	0	0	6.980	5.335
E097	t-cadinol	5.34	0.045	0.341	0.783	1	0	0	0	13.849	5.335
E098	alpha-santalene	5.93	0.045	0.388	0.571	1	0	0	0	22.453	5.345

Table E1. (Continued).

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
E099	citral	5.32	0.072	0.419	0.422	3	0	0	0	6.478	4.490
E100	geranial	5.08	0.065	0.446	0.333	3	0	0	0	6.478	4.705
E101	aromadendrene	5.53	0.043	0.337	0.620	0	0	0	0	22.453	5.455
E102	beta-selinene	5.89	0.053	0.364	0.816	0	0	0	0	21.446	5.470
E104	valencene	6.17	0.053	0.367	0.754	1	0	0	0	21.446	5.270
E105	fenchene	3.75	0.019	0.371	0.231	0	0	0	0	11.387	5.545
E106	geranyl formate	4.98	0.070	0.385	0.383	3	0	0	0	7.661	4.825
E107	(E),(E)-farnesol	6.13	0.094	0.374	0.706	3	0	0	0	15.283	5.070
E108	pregeijerene	5.86	0.135	0.367	0.571	4	0	0	0	14.008	5.195
E109	3,5- dimethoxytoluene	3.48	0.007	0.442	0.000	0	0	0	0	3.599	4.645
E110	3,4,5- trimethoxytoluene	3.23	0.016	0.400	-0.068	0	0	0	0	2.666	4.4900
E111	verbenone	4.04	0.016	0.382	0.295	1	0	0	0	5.062	4.870
E112	para-methoxycinnamic acid	4.52	0.031	0.508	-0.004	2	0	0	0	3.001	4.115
E113	2,2-dimethyl-6-vinylchroman-4-one	4.88	0.021	0.444	0.272	1	0	0	0	6.133	4.140
E114	2-senecioyl-4-vinylphenol	5.37	0.034	0.445	0.129	2	0	0	0	10.870	4.055
E115	trans-ethyl cinnamate	4.75	0.028	0.490	0.089	2	0	0	0	6.717	4.435
E116	hexyl butyrate	4.07	0.024	0.467	0.562	0	0	0	0	7.186	5.850
E117	benzyl salicylate	4.50	0.008	0.442	-0.015	0	0	0	0	11.951	4.340
E118	ethyl-p-methoxycinnamate	4.68	0.030	0.487	0.014	2	0	0	0	5.313	4.145
E119	alpha-cedrene	5.78	0.049	0.337	0.561	1	0	0	0	22.453	5.285
E120	beta-cedrene	5.52	0.044	0.339	0.626	0	0	0	0	22.453	5.490
E121	octyl acetate	4.57	0.024	0.463	0.530	0	0	0	0	12.907	5.835
E122	eucarvone	4.85	0.070	0.395	0.327	3	0	0	0	4.637	4.515
E123	bornyl acetate	3.64	0.014	0.377	0.356	0	0	0	0	8.186	5.610
E124	emodic acid	3.63	0.020	0.456	-0.167	0	0	0	0	0.940	3.745
E126	guineensine	7.89	0.304	0.338	0.643	6	0	0	3	16.484	4.110
E127	pipercide	7.86	0.292	0.476	0.553	6	0	0	3	13.207	4.135
E128	retrofractamide A	7.25	0.274	0.479	0.383	6	0	0	3	10.191	4.130
E129	(Z,Z)-matricaria ester	5.50	0.115	0.518	0.070	4	0	0	0	6.717	4.155
E130	(E)-cinnamic acid	4.45	0.028	0.501	0.035	2	0	0	0	3.987	4.400
E131	locustol	3.86	0.008	0.462	0.042	0	0	0	0	3.599	4.340
E132	coumestrol	3.57	0.022	0.469	-0.337	0	0	0	0	3.230	3.745
E133	parthenin	4.59	0.056	0.349	0.555	2	0	0	0	2.940	4.690
E134	(+)-camphene	3.79	0.021	0.351	0.300	0	0	0	0	11.387	5.565
E135	betulin	8.10	0.170	0.289	1.296	0	0	0	0	37.188	5.465
E136	quercetin	3.43	0.024	0.451	-0.225	0	0	0	0	0.054	3.725

Table E1. (Continued).

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
E137	parthenolide	4.86	0.040	0.355	0.614	1	0	0	0	6.826	4.710
E138	rutin	4.76	0.032	0.334	0.207	0	0	0	0	9.908	3.805
E139	enhydrin	4.21	0.051	0.325	0.593	1	0	0	0	1.669	4.710
E140	ferulic acid	4.33	0.034	0.492	-0.056	2	0	0	0	1.422	3.960
E141	(24R)-24,25-epoxycycloartan-3-one	8.31	0.160	0.347	1.194	0	0	0	0	37.188	5.190
E142	alantolactone	5.19	0.030	0.356	0.54	1	0	0	0	11.892	4.745
E143	isoalantolactone	5.17	0.038	0.361	0.661	0	0	0	0	11.892	4.780
E144	ergosterol endoperoxide	9.55	0.296	0.368	1.323	4	0	0	0	31.122	4.460
E145	pulegone	4.19	0.023	0.400	0.500	0	0	0	0	5.062	4.965
E146	apiole	4.12	0.018	0.425	0.083	1	0	0	0	2.525	4.105
E148	beta-pinene	4.20	0.016	0.370	0.435	0	0	0	0	11.387	5.455
E149	beta-eudesmol	4.95	0.040	0.339	0.775	0	0	0	0	13.849	5.505
E150	cis-carveol	3.99	0.014	0.409	0.384	1	0	0	0	5.062	5.290
E151	Z,E- nepetalactone	4.00	0.014	0.396	0.270	1	0	0	0	4.758	4.885
E152	E,Z- nepetalactone	3.92	0.013	0.394	0.220	1	0	0	0	4.758	4.860
E153	para-benzoquinone	4.68	0.113	0.471	0.098	4	0	0	0	0.169	4.250
E154	2-methyl parabenzoquinone	4.56	0.066	0.464	0.159	3	0	0	0	0.584	4.270
E155	2-isopropyl parabenzoquinone	4.76	0.069	0.431	0.234	3	0	0	0	2.001	4.225
M001	carvacryl glycolic acid	3.98	0.013	0.387	0.221	0	0	0	0	5.716	4.560
M002	1,8-cineole	3.06	0.025	0.346	0.365	0	0	0	0	6.262	6.015
M003	1,4-cineole	3.19	0.023	0.372	0.459	0	0	0	0	6.262	6.195
M004	carvacrol	4.39	0.010	0.416	0.268	0	0	0	0	7.914	4.560
M005	carvacryl benzoate	5.81	0.022	0.422	0.210	0	0	0	0	20.526	4.200
M006	carvacryl acetate	4.60	0.010	0.410	0.288	0	0	0	0	10.411	4.625
M007	carvacryl chloroacetate	4.94	0.010	0.438	0.262	0	0	0	0	12.239	4.470
M008	2-hydroxy-3-methyl-6,-(1-methylethyl)-benzaldehyde	4.92	0.019	0.406	0.231	1	0	0	0	8.483	4.105
M009	thymyl ethyl ether	4.63	0.012	0.383	0.263	0	0	0	0	11.463	4.530
M010	thymoxyacetic acid	4.14	0.015	0.388	0.233	0	0	0	0	5.716	4.390
M011	carvacryl propionate	4.99	0.014	0.418	0.353	0	0	0	0	12.239	4.555
M012	carvacryl trichloroacetate	5.34	0.013	0.451	0.168	0	0	0	0	16.145	4.275
M013	thymyl acetate	4.58	0.011	0.397	0.276	0	0	0	0	10.411	4.555
M014	thymyl chloroacetate	4.81	0.010	0.416	0.270	0	0	0	0	12.239	4.565
M015	thymyl trichloroacetate	5.39	0.014	0.444	0.206	0	0	0	0	16.145	4.280
M016	thymyl propionate	4.92	0.013	0.411	0.332	0	0	0	0	12.239	4.560
M017	thymyl benzoate	5.45	0.027	0.385	0.226	1	0	0	0	18.169	4.545

Table E1. (Continued).

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
M018	2-hydroxy-6-methyl-3-(1-methylethyl)-benzaldehyde	4.98	0.020	0.406	0.279	1	0	0	0	8.483	4.145
M019	5-norbornene-2-ol	3.12	0.045	0.403	0.001	2	0	0	0	1.969	5.340
M020	5-norbornene-2,2-dimethanol	3.11	0.046	0.373	0.095	2	0	0	0	1.377	5.385
M021	5-norbornene-2-endo-3-endodimethanol	3.24	0.044	0.362	0.181	2	0	0	0	1.377	5.385
M022	5-norbornene-2-exo-3-exo-dimethanol	3.31	0.040	0.375	0.191	2	0	0	0	1.377	5.370
M023	eugenyl acetate	4.50	0.010	0.452	0.148	1	0	0	0	6.620	4.425
M024	2-(2-methoxy-4-(2-propen-1-yl))-phenoxy acetic acid	4.01	0.010	0.456	0.074	1	0	0	0	3.116	4.440
M025	borneol	2.31	0.054	0.334	0.182	0	0	0	0	6.262	6.475
M026	catechol	3.39	0.010	0.472	-0.06	0	0	0	0	0.798	4.390
M027	alpha-terpinene	5.33	0.038	0.405	0.399	2	0	0	0	10.673	4.575
M028	terpineol	3.95	0.015	0.395	0.390	1	0	0	0	5.558	5.350
M029	1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	4.75	0.014	0.441	0.198	1	0	0	0	7.196	4.245
M030	eugenol	4.22	0.011	0.453	0.074	1	0	0	0	4.453	4.325
M031	phenol	3.45	0.007	0.477	-0.019	0	0	0	0	2.268	4.615
M032	g-terpinene	4.92	0.036	0.402	0.416	2	0	0	0	10.673	5.145
M033	guaiacol	3.46	0.009	0.469	-0.058	0	0	0	0	1.553	4.375
M034	1-benzoate-2-methoxy-4-(3-hydroxypropyl)-phenol	4.50	0.011	0.423	-0.01	0	0	0	0	9.954	4.005
M035	4-hydroxy-3-methoxy-benzenepropanol	3.71	0.017	0.407	0.145	0	0	0	0	1.862	4.315
M036	isoborneol	2.24	0.058	0.330	0.153	0	0	0	0	6.262	6.470
M037	isopulegol	3.95	0.023	0.389	0.580	0	0	0	0	5.558	5.495
M038	thymol	4.35	0.009	0.417	0.255	0	0	0	0	7.914	4.575
M039	menthone	4.12	0.025	0.389	0.572	0	0	0	0	5.558	5.245
M040	nonan-2-one	4.92	0.031	0.476	0.615	0	0	0	0	11.277	5.435
M041	undecan-2-one	5.65	0.054	0.474	0.773	0	0	0	0	15.626	5.435
M042	1,2-dimethoxy-4-(2-propen-1-yl)-benzene	4.51	0.011	0.457	0.125	1	0	0	0	5.768	4.265
M043	neo-isopulegol	3.83	0.018	0.386	0.513	0	0	0	0	5.558	5.465
M044	1,2-carvone oxide	3.81	0.023	0.422	0.454	0	0	0	0	1.875	5.040
M045	limonene oxide,cis	3.74	0.014	0.407	0.445	0	0	0	0	5.558	5.510
M047	p-cymene	4.75	0.009	0.419	0.278	0	0	0	0	12.69	4.735
M048	eugenyl propionate	4.56	0.009	0.452	0.091	1	0	0	0	8.093	4.390
M049	R-carvone	4.25	0.015	0.421	0.358	1	0	0	0	4.637	4.885
M050	S-carvone	4.13	0.012	0.421	0.299	1	0	0	0	4.637	4.890

Table E1. (Continued)

Label	Chemical	pEC_{50}	Hat value (*h**= 0.069)	SPAM	Mor31p	NdsCH	CATS2D_02_AP	B05 [C-S]	F03 [C-N]	MLOGP2	Hardness
M051	R-limonene	4.60	0.016	0.405	0.408	1	0	0	0	10.673	5.250
M052	S-limonene	4	0.016	0.405	0.406	1	0	0	0	10.673	5.250
M053	resorcinol	3.27	0.009	0.478	-0.044	0	0	0	0	0.798	4.605
M054	salicyl aldehyde	4.01	0.012	0.479	-0.031	1	0	0	0	2.779	4.240
M055	vanillin	3.75	0.016	0.461	-0.086	1	0	0	0	0.850	4.120
M056	2,6-dimethyl-p-benzoquinone	4.39	0.036	0.440	0.212	2	0	0	0	1.205	4.280
M057	2,5-dimethyl-p-benzoquinone	4.46	0.037	0.444	0.240	2	0	0	0	1.205	4.290
M058	thymoquinone	4.74	0.042	0.419	0.325	2	0	0	0	2.952	4.245
M059	pipilyasine	8.95	0.232	0.474	0.974	4	0	0	2	25.773	4.550
M060	pipzubedine	9.88	0.287	0.346	1.190	4	0	0	2	35.74	4.570
M061	pipyaqubine	9.71	0.280	0.478	1.096	4	0	0	1	31.207	4.375
M062	pellitorine	7.19	0.165	0.475	0.763	4	0	0	2	11.796	4.540
M063	pipericine	9.08	0.218	0.339	1.434	0	0	0	3	38.223	5.470
M064	piperine	6.09	0.119	0.448	0.269	4	0	0	2	6.629	3.950
M065	(-)-camphene	3.79	0.021	0.351	0.300	0	0	0	0	11.387	5.565
M066	3-carene	4.62	0.019	0.387	0.449	1	0	0	0	11.387	5.335
M067	camphor	3.38	0.018	0.344	0.263	0	0	0	0	5.558	5.230
M068	menthol	3.20	0.029	0.375	0.570	0	0	0	0	6.262	6.475
M069	tetradecanoic acid	5.60	0.049	0.342	0.744	0	0	0	0	22.095	5.655
M070	2,4-di-tert-butylphenol	5.12	0.020	0.359	0.339	0	0	0	0	15.375	4.480
M071	linoleic acid	7.36	0.191	0.33	1.035	4	0	0	0	20.891	5.110
M072	nerolidol	5.94	0.097	0.33	0.696	3	0	0	0	15.283	5.090
M073	palmitic acid	6.71	0.091	0.369	1.019	0	0	0	0	27.103	5.650
M074	methyl linolelaidate	7.41	0.187	0.333	0.956	4	0	0	0	23.069	5.145
M075	beta-caryophyllene	6.00	0.051	0.335	0.684	1	0	0	0	21.446	5.175
M076	geranic acid	4.78	0.030	0.453	0.302	2	0	0	0	6.133	4.690
M077	terpinen-4-ol	4.11	0.020	0.384	0.486	1	0	0	0	5.558	5.335
M078	ethyl palmitate	7.66	0.152	0.469	1.141	0	0	0	0	32.341	5.84
M079	humulene	6.45	0.160	0.336	0.609	4	0	0	0	20.56	5.215
M080	behenic acid	8.66	0.195	0.35	1.227	0	0	0	0	43.408	5.625
M081	n-hexadecane	7.49	0.232	0.456	1.197	0	0	0	0	41.433	7.315
M082	trans-anethole	4.84	0.029	0.482	0.073	2	0	0	0	7.414	4.34
M083	estragole	4.61	0.009	0.47	0.145	1	0	0	0	7.414	4.455

Table E2: Chemicals with their labels, their predicted pLC_{50} values from Eq. 4.10, hat values and descriptor values

Label	Chemical Name	pLC_{50} (µmol/L)	Hat values (h^*=0.692)	nRCOOH	E_{HOMO} (eV)
E001	g-elemene	-1.12	0.144	0	-8.80
E002	1-(cyclohexylacetyl)-2-methyl-piperidine	-1.29	0.115	0	-9.16
E003	(2R)-1-decanoyl-2-methyl-piperidine	-1.29	0.114	0	-9.18
E004	1-dodecanoyl-2-methyl-piperidine	-1.28	0.115	0	-9.15
E005	(2R)-1-heptanoyl-2-methyl-piperidine	-1.29	0.114	0	-9.18
E006	1-(3-cyclohexylpropanoyl)-2-methyl-piperidine	-1.28	0.115	0	-9.15
E007	1-[(4-methylcyclohexyl)carbonyl]-2-methyl-piperidine	-1.28	0.116	0	-9.14
E008	(3S)-1-(1-methylcyclohexyl)carbonyl-3-methyl-piperidine	-1.24	0.120	0	-9.07
E009	(3S)-1-(3-cyclohexylpropanoyl)-3-methyl-piperidine	-1.27	0.116	0	-9.13
E010	(3S)-1-heptanoyl-3-methyl-piperidine	-1.28	0.116	0	-9.14
E011	(3S)-1-(cyclohexylcarbonyl)-3-methyl-piperidine	-1.27	0.117	0	-9.12
E012	1-decanoyl-4-methyl-piperidine	-1.27	0.116	0	-9.13
E013	1-(4-cyclohexylbutanoyl)-4-methyl-piperidine	-1.29	0.114	0	-9.17
E014	1-(cyclohexylcarbonyl)-4-methyl-piperidine	-1.28	0.116	0	-9.14
E015	1-(3-cyclohexylpropanoyl)-4-methyl-piperidine	-1.29	0.114	0	-9.17
E016	1-dodecanoyl-4-methyl-piperidine	-1.29	0.114	0	-9.18
E017	1-(cyclohexylcarbonyl)-2-ethyl-piperidine	-1.25	0.119	0	-9.08
E018	1-(3-cyclohexylpropanoyl)-2-ethyl-piperidine	-1.26	0.117	0	-9.11
E019	1-propionyl-2-ethyl-piperidine	-1.27	0.117	0	-9.12
E020	1-(3-cyclopentylpropanoyl)-2-ethyl-piperidine	-1.28	0.116	0	-9.14
E021	1-nonanoyl-2-ethyl-piperidine	-1.27	0.116	0	-9.13
E022	1-octanoyl-3-benzyl-piperidine	-1.28	0.116	0	-9.14
E023	1-undec-10-enoyl-4-benzyl-piperidine	-1.29	0.114	0	-9.18
E024	1-cyclohexylacetyl-4-benzyl-piperidine	-1.30	0.113	0	-9.20
E025	1-(3-cyclohexylpropanoyl)-4-benzyl-piperidine	-1.30	0.113	0	-9.20
E026	2-methyl-1-undec-10-enoyl-piperidine	-1.30	0.113	0	-9.20
E027	2-ethyl-1-undec-10-enoyl-piperidine	-1.27	0.116	0	-9.13
E028	2-benzyl-1-undec-10-enoyl-piperidine	-1.29	0.114	0	-9.17
E029	3-methyl-1-undec-10-enoyl-piperidine	-1.28	0.115	0	-9.15
E030	3-ethyl-1-undec-10-enoyl-piperidine	-1.28	0.115	0	-9.15
E031	3-benzyl-1-undec-10-enoyl-piperidine	-1.28	0.115	0	-9.15
E032	4-methyl-1-undec-10-enoyl-piperidine	-1.31	0.113	0	-9.21
E033	4-ethyl-1-undec-10-enoyl-piperidine	-1.29	0.114	0	-9.18
E035	alpha-pinene	-1.19	0.128	0	-8.96
E037	safrole	-1.07	0.161	0	-8.68
E038	piperitone	-1.50	0.119	0	-9.63
E040	trans-ocimenone	-1.48	0.116	0	-9.58
E041	terpinolene	-1.08	0.156	0	-8.71
E042	alpha-bisabolol	-1.26	0.118	0	-9.10
E043	alpha-cadinol	-1.24	0.121	0	-9.05
E044	t-muurolol	-1.30	0.114	0	-9.19
E045	cis-isolongifolone	-1.49	0.118	0	-9.61
E046	delta-cadinene	-1.08	0.158	0	-8.70
E047	tectoquinone	-1.63	0.146	0	-9.93
E048	guaiol	-1.09	0.153	0	-8.73
E049	citronellic acid	-2.41	0.251	1	-9.22
E050	myrtenol	-1.41	0.111	0	-9.44

Table E2. (Continued).

Label	Chemical Name	pLC_{50} (µmol/L)	Hat values (h^*=0.692)	nRCOOH	E_{HOMO} (eV)
E051	16-kaurene	-1.44	0.113	0	-9.51
E052	elemol	-1.47	0.115	0	-9.56
E053	cedrol	-1.64	0.149	0	-9.95
E054	epi-zonarene	-0.83	0.267	0	-8.16
E055	beta-guaiene	-1.01	0.180	0	-8.56
E056	ascaridole	-1.04	0.171	0	-8.61
E057	spathulenol	-1.44	0.113	0	-9.50
E058	p-anisaldehyde	-1.35	0.111	0	-9.31
E059	m-eugenol	-0.99	0.189	0	-8.51
E060	germacrene D	-1.24	0.120	0	-9.06
E061	benzyl benzoate	-1.50	0.119	0	-9.63
E062	methyl-cinnamate	-1.51	0.120	0	-9.65
E063	piperitone oxide	-1.56	0.129	0	-9.77
E064	fenchone	-1.36	0.111	0	-9.33
E065	cinnamaldehyde	-1.48	0.117	0	-9.60
E066	alpha-phellandrene	-1.03	0.175	0	-8.59
E067	cinnamyl acetate	-1.34	0.111	0	-9.29
E068	beta-phellandrene	-1.21	0.125	0	-9.00
E069	linalool	-1.30	0.113	0	-9.20
E070	caryophyllene epoxide	-1.51	0.121	0	-9.66
E071	alpha-eudesmol	-1.16	0.134	0	-8.89
E072	p-menthane-3,8-diol	-1.67	0.157	0	-10.01
E073	citronellal	-1.37	0.111	0	-9.34
E074	myristicin	-1.05	0.165	0	-8.65
E075	dillapiole	-1.00	0.185	0	-8.53
E076	alpha-copaene	-1.17	0.132	0	-8.91
E077	asaricin	-0.85	0.256	0	-8.20
E078	1-butyl-3,4-methylenedioxybenzene	-1.04	0.170	0	-8.62
E079	isoelemicin	-0.99	0.191	0	-8.50
E080	Z-asarone	-0.83	0.267	0	-8.16
E081	patchouli alcohol	-1.51	0.120	0	-9.65
E082	alpha-asarone	-0.70	0.354	0	-7.86
E083	geijerene	-1.44	0.113	0	-9.51
E084	sabinene	-1.42	0.112	0	-9.45
E085	viridiflorol	-1.61	0.140	0	-9.88
E086	bicyclogermacrene	-1.00	0.183	0	-8.54
E088	curcumene	-1.16	0.134	0	-8.89
E089	ar-turmerone	-1.22	0.123	0	-9.02
E090	zingiberene	-1.02	0.178	0	-8.57
E091	beta-turmerone	-1.24	0.121	0	-9.05
E092	dodecanal	-1.69	0.164	0	-10.06
E093	1-dodecanol	-1.86	0.236	0	-10.43
E094	(E)-beta-ocimene	-1.22	0.124	0	-9.01
E095	myrcene epoxide	-1.48	0.116	0	-9.58
E096	dihydrotagetone	-1.63	0.144	0	-9.91
E097	t-cadinol	-1.28	0.115	0	-9.15
E098	alpha-santalene	-1.26	0.117	0	-9.11
E099	neral	-1.24	0.120	0	-9.06
E100	geranial	-1.41	0.111	0	-9.43
E101	aromadendrene	-1.41	0.111	0	-9.44
E102	beta-selinene	-1.44	0.113	0	-9.51
E104	valencene	-1.26	0.118	0	-9.10
E105	fenchene	-1.47	0.115	0	-9.56
E106	geranyl formate!	-1.25	0.119	0	-9.08
E107	(E),(E)-farnesol	-1.23	0.122	0	-9.03
E108	pregeijerene	-1.21	0.125	0	-9.00

Table E2. (Continued).

Label	Chemical Name	pLC_{50} (µmol/L)	Hat values (*h**=0.692)	nRCOOH	E_{HOMO} (eV)
E109	3,5- dimethoxytoluene	-1.09	0.155	0	-8.72
E110	3,4,5- trimethoxytoluene	-1.01	0.182	0	-8.55
E111	verbenone	-1.54	0.126	0	-9.73
E112	para-methoxycinnamic acid	-2.34	0.250	1	-9.07
E113	2,2-dimethyl-6-vinylchroman-4-one	-1.15	0.138	0	-8.86
E114	2-senecioyl-4-vinylphenol	-1.12	0.144	0	-8.80
E115	trans-ethyl cinnamate	-1.49	0.118	0	-9.62
E116	hexyl butyrate	-2.05	0.349	0	-10.84
E117	benzyl salicylate	-1.39	0.111	0	-9.39
E118	ethyl-p-methoxycinnamate	-1.19	0.128	0	-8.96
E119	alpha-cedrene	-1.18	0.131	0	-8.93
E120	beta-cedrene	-1.43	0.112	0	-9.47
E121	octyl acetate	-2.04	0.343	0	-10.82
E122	eucarvone	-1.36	0.111	0	-9.33
E123	bornyl acetate	-1.79	0.203	0	-10.28
E124	emodic acid	-1.58	0.134	0	-9.82
E126	guineensine	-0.95	0.208	0	-8.41
E127	pipercide	-1.05	0.168	0	-8.63
E128	retrofractamide A	-1.06	0.162	0	-8.67
E129	(Z,Z)-matricaria ester	-1.33	0.112	0	-9.25
E130	(E)-cinnamic acid	-2.65	0.291	1	-9.74
E131	locustol	-0.96	0.200	0	-8.45
E132	coumestrol	-1.14	0.141	0	-8.83
E133	parthenin	-1.70	0.166	0	-10.07
E134	(+)-camphene	-1.51	0.121	0	-9.66
E135	betulin	-1.51	0.120	0	-9.65
E136	quercetin	-1.12	0.144	0	-8.80
E137	parthenolide	-1.56	0.129	0	-9.77
E138	rutin	-1.21	0.125	0	-8.99
E139	enhydrin	-1.79	0.203	0	-10.28
E140	ferulic acid	-2.26	0.256	1	-8.89
E141	(24R)-24,25-epoxycycloartan-3-one	-1.48	0.117	0	-9.60
E142	alantolactone	-1.53	0.123	0	-9.69
E143	isoalantolactone	-1.58	0.134	0	-9.82
E144	ergosterol endoperoxide	-1.07	0.161	0	-8.68
E145	pulegone	-1.44	0.113	0	-9.50
E146	apiole	-0.89	0.237	0	-8.28
E148	beta-pinene	-1.42	0.112	0	-9.46
E149	beta-eudesmol	-1.44	0.113	0	-9.51
E150	cis-carveol	-1.46	0.115	0	-9.55
E151	Z,E nepetalactone	-1.40	0.111	0	-9.42
E152	E,Z nepetalactone	-1.39	0.111	0	-9.40
E153	para-benzoquinone	-1.93	0.276	0	-10.59
E154	2-methyl parabenzoquinone	-1.88	0.247	0	-10.48
E155	2-isopropyl parabenzoquinone	-1.83	0.220	0	-10.36
M001	carvacryl glycolic acid	-2.40	0.250	1	-9.20
M002	1,8-cineole	-1.38	0.111	0	-9.37
M003	1,4-cineole	-1.53	0.124	0	-9.71
M004	carvacrol	-1.10	0.152	0	-8.74
M005	carvacryl benzoate	-1.24	0.120	0	-9.06
M006	carvacryl acetate	-1.25	0.119	0	-9.08
M007	carvacryl chloroacetate	-1.30	0.113	0	-9.20
M008	2-hydroxy-3-methyl-6-(1-methylethyl)-benzaldehyde	-1.19	0.128	0	-8.96
M009	thymyl ethyl ether	-0.98	0.192	0	-8.49
M010	thymoxyacetic acid	-2.22	0.261	1	-8.80

Table E2. (Continued).

Label	Chemical Name	pLC_{50} (μmol/L)	Hat values (h^*=0.692)	nRCOOH	E_{HOMO} (eV)
M011	carvacryl propionate	-1.17	0.132	0	-8.91
M012	carvacryl trichloroacetate	-1.41	0.111	0	-9.44
M013	thymyl acetate	-1.19	0.129	0	-8.95
M014	thymyl chloroacetate	-1.34	0.111	0	-9.27
M015	thymyl trichloroacetate	-1.41	0.111	0	-9.44
M016	thymyl propionate	-1.18	0.131	0	-8.92
M017	thymyl benzoate	-1.23	0.122	0	-9.03
M018	2-hydroxy-6-methyl-3-(1-methylethyl)-benzaldehyde	-1.23	0.122	0	-9.03
M019	5-norbornene-2-ol	-1.56	0.128	0	-9.76
M020	5-norbornene-2,2-dimethanol	-1.63	0.146	0	-9.93
M021	5-norbornene-2-endo-3-endodimethanol	-1.66	0.154	0	-9.99
M022	5-norbornene-2-exo-3-exo-dimethanol	-1.63	0.144	0	-9.91
M023	eugenyl acetate	-1.14	0.141	0	-8.83
M024	2-(2-methoxy-4-(2-propen-1-yl))phenoxy acetic acid	-2.44	0.252	1	-9.28
M025	borneol	-1.64	0.147	0	-9.94
M026	catechol	-1.13	0.142	0	-8.82
M027	alpha-terpinene	-0.91	0.228	0	-8.32
M028	terpineol	-1.24	0.120	0	-9.06
M029	1-ethoxy-2-methoxy-4-(2-propen-1-yl) benzene	-0.85	0.259	0	-8.19
M030	eugenol	-0.99	0.191	0	-8.50
M031	phenol	-1.26	0.118	0	-9.10
M032	g-terpinene	-1.10	0.149	0	-8.76
M033	guaiacol	-1.04	0.170	0	-8.62
M034	1-benzoate-2-methoxy-4-(3-hydroxypropyl)-phenol	-1.15	0.137	0	-8.87
M035	4-hydroxy-3-methoxy-benzenepropanol	-1.04	0.170	0	-8.62
M036	isoborneol	-1.64	0.147	0	-9.94
M037	isopulegol	-1.47	0.116	0	-9.57
M038	thymol	-1.10	0.151	0	-8.75
M039	menthone	-1.49	0.118	0	-9.62
M040	nonan-2-one	-1.72	0.174	0	-10.12
M041	undecan-2-one	-1.72	0.174	0	-10.12
M042	1,2-dimethoxy-4-(2-propen-1-yl)-benzene	-0.88	0.242	0	-8.26
M043	neo-isopulegol	-1.46	0.114	0	-9.54
M044	1,2-carvone oxide	-1.67	0.158	0	-10.02
M045	limonene oxide, cis	-1.56	0.129	0	-9.77
M047	p-cymene	-1.18	0.131	0	-8.92
M048	eugenyl propionate	-1.10	0.151	0	-8.75
M049	R-carvone	-1.60	0.137	0	-9.85
M050	S-carvone	-1.60	0.137	0	-9.85
M051	R-limonene	-1.25	0.119	0	-9.08
M052	S-limonene	-1.25	0.119	0	-9.08
M053	resorcinol	-1.28	0.116	0	-9.14
M054	salicyl aldehyde	-1.41	0.111	0	-9.43
M055	vanillin	-1.22	0.123	0	-9.02
M056	2,6-dimethyl-para-benzoquinone	-1.83	0.222	0	-10.37
M057	2,5-dimethyl-para-benzoquinone	-1.83	0.222	0	-10.37
M058	thymoquinone	-1.78	0.197	0	-10.25
M059	pipilyasine	-1.39	0.111	0	-9.40
M060	pipzubedine	-1.36	0.111	0	-9.33
M061	pipyaqubine	-1.21	0.125	0	-8.99
M062	pellitorine	-1.39	0.111	0	-9.39
M063	pipericine	-1.49	0.118	0	-9.62

Table E2. (Continued).

Label	Chemical Name	pLC_{50} (μmol/L)	Hat values (*h**=0.692)	nRCOOH	E_{HOMO} (eV)
M064	piperine	-1.10	0.152	0	-8.74
M065	(-)-camphene	-1.51	0.121	0	-9.66
M066	3-carene	-1.22	0.123	0	-9.02
M067	camphor	-1.48	0.116	0	-9.58
M068	menthol	-1.71	0.169	0	-10.09
M069	tetradecanoic acid	-3.04	0.492	1	-10.62
M070	2,4-di-tert-butylphenol	-1.01	0.182	0	-8.55
M071	linoleic acid	-2.54	0.265	1	-9.50
M072	nerolidol	-1.22	0.123	0	-9.02
M073	palmitic acid	-3.04	0.489	1	-10.61
M074	methyl linolelaidate	-1.42	0.112	0	-9.46
M075	beta-caryophyllene	-1.18	0.131	0	-8.92
M076	geranic acid	-2.54	0.265	1	-9.50
M077	terpinen-4-ol	-1.34	0.111	0	-9.29
M078	ethyl palmitate	-2.01	0.327	0	-10.77
M079	humulene	-1.17	0.132	0	-8.91
M080	behenic acid	-3.02	0.476	1	-10.57
M081	n-hexadecane	-1.99	0.312	0	-10.72
M082	trans-anethole	-0.95	0.208	0	-8.41
M083	estragole	-1.05	0.167	0	-8.64

Table E3: Chemicals with their labels, their predicted pLC_{50} values from Eq. 4.11, hat values and descriptor values.

Label	Chemical name	pLC_{50} (mol/L)	Hat values (h^*=0.383)	logK_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s
E035	alpha-pinene	-2.57	0.309	4.48	0.000	0.968	0	1.333
E037	safrole	-1.88	0.068	3.45	0.725	1.013	0	0.385
E038	piperitone	-1.02	0.211	2.85	1.672	0.968	0	2.093
E041	terpinolene	-0.27	0.299	4.47	2.031	0.966	0	1.457
E042	alpha-bisabolol	-1.07	0.212	5.63	1.342	0.895	0	2.284
E045	cis-isolongifolone	-2.73	0.286	3.81	0.000	0.967	0	1.585
E046	delta-cadinene	-0.49	0.150	6.32	1.235	0.961	0	2.808
E047	tectoquinone	-0.89	0.176	3.89	1.188	1.058	0	0.707
E048	guaiol	-1.07	0.121	5.01	1.063	0.958	0	2.816
E049	citronellic acid	-1.58	0.202	3.78	1.474	0.888	2	0.121
E050	myrtenol	-3.09	0.251	3.22	0.000	0.959	0	0.604
E052	elemol	-0.83	0.149	5.54	1.277	0.937	0	2.759
E053	cedrol	-1.88	0.114	4.33	0.583	0.967	0	1.863
E055	beta-guaiene	-0.56	0.164	6.56	1.118	0.958	0	2.918
E056	ascaridole	-0.79	0.191	3.57	1.718	0.974	0	1.653
E058	p-anisaldehyde	-1.25	0.328	1.76	1.907	0.992	0	-0.274
E061	benzyl benzoate	-1.88	0.078	3.97	1.013	0.969	0	-0.837
E062	methyl-cinnamate	-1.77	0.089	2.62	1.176	1.011	0	-0.648
E063	piperitone oxide	-1.03	0.197	2.89	1.718	0.979	0	1.126
E064	fenchone	-3.01	0.232	3.04	0.000	0.981	0	0.515
E065	cinnamaldehyde	-2.44	0.056	1.9	0.678	1.011	0	-0.007
E067	cinnamyl acetate	-2.24	0.042	2.85	0.832	0.982	0	-0.334
E068	beta-phellandrene	-0.64	0.162	4.7	1.693	0.963	0	1.259
E069	linalool	-1.42	0.177	2.97	1.530	0.932	0	1.838
E070	caryophyllene epoxide	-1.16	0.144	5.25	1.038	0.938	0	2.854
E071	alpha-eudesmol	-0.70	0.157	4.81	1.496	0.952	0	2.655
E072	p-menthane-3,8-diol	-1.85	0.106	2.29	1.302	0.948	0	1.291
E073	citronellal	-1.20	0.133	3.53	1.511	0.950	0	1.613
E074	myristicin	-2.21	0.078	3.53	0.562	0.994	0	0.114
E075	dillapiole	-2.18	0.059	3.61	0.688	0.975	0	0.011
E076	alpha-copaene	-0.90	0.102	5.36	1.116	0.970	0	2.336
E077	asaricin	-1.78	0.041	3.53	1.027	0.978	0	0.181
E079	isoelemicin	-2.53	0.058	2.82	0.744	0.957	0	-0.461
E080	Z-asarone	-2.44	0.080	3.03	0.917	0.932	0	-0.445
E081	patchouli alcohol	-1.85	0.075	3.98	0.829	0.950	0	1.636
E082	alpha-asarone	-2.26	0.050	3.03	0.917	0.954	0	-0.215
E084	sabinene	-2.41	0.317	4.69	0.000	0.982	0	1.421
E085	viridiflorol	-2.05	0.235	4.63	0.335	0.954	0	2.653
E088	curcumene	-0.84	0.152	6.29	1.245	0.926	0	1.982
E090	zingiberene	-0.49	0.167	6.92	1.286	0.939	0	2.261
E092	dodecanal	-1.22	0.096	4.75	0.925	0.980	0	2.363
E093	1-dodecanol	-0.71	0.137	5.13	0.936	1.018	1	2.531
E094	(E)-beta-ocimene	-0.68	0.184	4.8	1.719	0.942	0	1.548
E095	myrcene epoxide	-1.02	0.151	3.48	1.610	0.966	0	1.444

Table E3. (Continued).

Label	Chemical name	pLC_{50} (mol/L)	Hat values (h^*=0.383)	logK_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s
E096	dihydrotagetone	-1.64	0.113	2.92	1.273	0.943	0	1.933
E098	alpha-santalene	-0.47	0.139	6.43	1.334	0.956	0	2.189
E099	neral	-1.15	0.237	3.45	1.751	0.923	0	1.499
E100	geranial	-0.99	0.207	3.45	1.751	0.948	0	1.465
E102	beta-selinene	-0.08	0.195	6.3	1.583	0.959	0	2.891
E104	valencene	-0.80	0.159	6.3	0.947	0.964	0	2.795
E105	fenchene	-2.59	0.294	4.35	0.000	0.977	0	1.075
E106	geranyl formate	-1.96	0.295	3.93	1.336	0.859	0	1.153
E107	(E),(E)-farnesol	-0.78	0.382	5.77	1.537	0.851	2	2.978
E109	3,5-dimethoxytoluene	-2.79	0.091	2.70	0.414	0.973	0	-0.196
E111	verbenone	-2.98	0.240	3.21	0.000	0.976	0	0.619
E112	para-methoxycinnamic acid	-2.04	0.061	2.68	1.010	0.998	0	-0.743
E115	trans-ethyl cinnamate	-1.91	0.044	2.99	0.971	0.996	0	-0.083
E116	hexyl butyrate	-1.77	0.055	3.81	0.808	0.978	0	1.371
E117	benzyl salicylate	-1.63	0.086	4.31	0.973	0.966	2	-0.849
E118	ethyl-p-methoxycinnamate	-1.95	0.065	2.93	1.097	0.987	0	-0.777
E119	alpha-cedrene	-1.36	0.176	5.74	0.596	0.975	0	2.310
E120	beta-cedrene	-1.35	0.190	5.82	0.561	0.978	0	2.348
E121	octyl acetate	-1.86	0.052	3.81	0.898	0.956	0	1.056
E122	eucarvone	-2.98	0.233	2.89	0.000	0.981	0	1.001
E123	bornyl acetate	-1.85	0.056	3.86	0.861	0.958	0	1.245
E124	emodic acid	-1.15	0.138	3.34	1.177	1.047	1	-0.182
E130	(E)-cinnamic acid	-2.32	0.060	2.13	0.762	1.014	0	-0.412
E131	locustol	-2.35	0.036	2.38	0.869	0.972	0	-0.079
E132	coumestrol	-1.39	0.321	1.57	1.199	1.067	4	-1.762
E133	parthenin	-2.44	0.106	0.77	0.98	0.972	0	1.531
E134	(+)-camphene	-2.60	0.299	4.35	0.000	0.971	0	1.258
E136	quercetin	-2.02	0.262	1.48	1.069	1.019	4	-2.784
E138	rutin	-2.52	1.387	-1.11	1.092	0.827	12	1.230
E140	ferulic acid	-2.78	0.079	1.42	0.786	0.994	0	-1.504
E142	alantolactone	-1.59	0.055	3.38	1.189	0.966	0	1.007
E143	isoalantolactone	-1.39	0.067	3.42	1.283	0.978	0	1.056
E144	ergosterol endoperoxide	0.44	0.334	8.71	1.063	0.973	2	3.921
E145	pulegone	-1.29	0.122	3.08	1.393	0.967	0	2.043
E146	apiole	-2.29	0.073	3.61	0.584	0.971	0	0.244
E148	beta-pinene	-2.66	0.300	4.16	0.000	0.964	0	1.514
E149	beta-eudesmol	-0.72	0.146	4.88	1.496	0.951	0	2.405
E153	para-benzoquinone	-3.41	0.236	0.2	0.000	1.035	0	0.733
E154	2-methyl-para-benzoquinone	-3.28	0.233	0.72	0.000	1.024	0	1.156
M002	1,8-cineole	-3.12	0.248	2.74	0.000	0.960	0	1.258
M003	1,4-cineole	-2.91	0.255	2.97	0.000	0.974	0	1.599
M004	carvacrol	-1.52	0.051	3.49	1.159	0.976	0	1.060

Table E3. (Continued).

Label	Chemical name	pLC_{50} (mol/L)	Hat values (h^*=0.383)	logK_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s
M006	carvacryl acetate	-1.57	0.054	3.59	1.209	0.969	0	0.452
M010	thymoxyacetic acid	-2.10	0.073	3.33	1.080	0.939	0	-0.279
M013	thymyl acetate	-1.52	0.082	3.59	1.328	0.948	0	0.850
M016	thymyl propionate	-1.28	0.075	4.08	1.312	0.959	0	1.193
M019	5-norbornene-2-ol	-3.67	0.192	0.99	0.000	0.996	0	-0.752
M023	eugenyl acetate	-2.19	0.039	3.06	0.856	0.962	0	0.273
M025	borneol	-3.23	0.229	2.69	0.000	0.964	0	0.400
M026	catechol	-3.69	0.192	0.88	0.000	0.998	0	-0.772
M027	alpha-terpinene	-0.75	0.150	4.25	1.625	0.968	0	1.648
M028	terpineol	-1.07	0.210	2.98	1.760	0.962	0	1.108
M030	eugenol	-2.69	0.051	2.27	0.635	0.963	0	0.086
M031	phenol	-3.42	0.193	1.46	0.000	1.000	0	-0.169
M032	g-terpinene	-0.65	0.150	4.50	1.625	0.971	0	1.735
M033	guaiacol	-3.68	0.197	1.32	0.000	0.977	0	-0.643
M035	4-hydroxy-3-methoxy-benzenepropanol	-3.12	0.101	1.40	0.732	0.925	0	-0.285
M036	isoborneol	-3.06	0.248	3.24	0.000	0.963	0	0.588
M037	isopulegol	-1.18	0.132	3.37	1.485	0.960	0	1.864
M038	thymol	-1.09	0.153	3.30	1.622	0.967	0	1.226
M039	menthone	-1.64	0.081	2.87	1.234	0.960	0	1.573
M040	nonan-2-one	-1.70	0.069	3.14	0.867	0.997	0	1.837
M041	undecan-2-one	-1.19	0.115	4.09	0.912	1.012	0	2.444
M042	1,2-dimethoxy-4-(2-propen-1-yl)-benzene	-2.38	0.053	3.03	0.654	0.966	0	0.427
M044	1,2-carvone oxide	-0.49	0.388	2.88	2.084	0.986	0	1.668
M047	p-cymene	-0.85	0.131	4.10	1.577	0.975	0	1.328
M048	eugenyl propionate	-2.17	0.068	3.55	0.899	0.959	0	-0.718
M049	R-carvone	-0.65	0.357	2.71	2.028	0.982	0	1.492
M050	S-carvone	-0.60	0.336	3.07	2.028	0.982	0	1.196
M051	R-limonene	-0.17	0.302	4.57	2.031	0.973	0	1.621
M052	S-limonene	-0.17	0.302	4.57	2.031	0.973	0	1.613
M053	resorcinol	-3.67	0.189	0.80	0.000	1.000	0	-0.596
M054	salicyl aldehyde	-3.18	0.206	1.81	0.000	1.006	0	0.566
M055	vanillin	-3.72	0.200	1.21	0.011	0.986	0	-1.147
M056	2,6-dimethyl-para-benzoquinone	-3.09	0.253	1.22	0.000	1.019	0	1.722
M057	2,5-dimethyl-para-benzoquinone	-3.06	0.255	1.28	0.000	1.019	0	1.762
M058	thymoquinone	-1.19	0.204	2.20	1.625	0.989	0	1.616
M062	pellitorine	-0.85	0.114	4.20	1.059	1.016	1	2.507
M064	piperine	-1.42	0.068	3.69	1.041	1.015	0	0.567
M065	(-)-camphene	-2.60	0.299	4.35	0.000	0.971	0	1.257
M066	3-carene	-2.44	0.312	4.38	0.000	0.982	0	1.809
M067	camphor	-3.15	0.221	2.38	0.000	0.977	0	0.946
M068	menthol	-1.54	0.080	3.40	1.277	0.951	0	1.35
M069	tetradecanoic acid	-1.79	0.452	6.11	0.947	0.829	1	1.727

Table E3. (Continued).

Label	Chemical name	pLC_{50} (mol/L)	Hat values (h^*=0.383)	logK_{ow}	GATS7p	SpMaxA_G/D	CATS2D_08_DL	Mor31s
M070	2,4-di-tert-butylphenol	-0.73	0.141	5.19	1.589	0.948	0	1.181
M071	linoleic acid	-1.59	0.738	7.05	1.004	0.787	1	2.778
M072	nerolidol	-1.23	0.492	5.68	1.504	0.821	1	2.432
M073	palmitic acid	-1.36	0.507	7.17	0.957	0.830	1	2.517
M074	methyl linolelaidate	-1.33	0.562	7.80	1.003	0.825	0	2.15
M075	beta-caryophyllene	-0.86	0.166	6.30	1.128	0.929	0	2.584
M076	geranic acid	-0.93	0.145	3.70	1.673	0.951	2	0.376
M077	terpinen-4-ol	-1.28	0.125	3.26	1.44	0.957	0	1.859
M078	ethyl palmitate	0.07	0.302	7.74	0.965	1.025	0	3.186
M079	humulene	-0.80	0.212	6.95	1.146	0.912	0	2.399
M081	n-hexadecane	0.46	0.443	8.20	0.955	1.04	0	4.384
M082	trans-anethole	-1.79	0.038	3.39	0.963	0.987	0	0.435
M083	estragole	-1.66	0.043	3.47	1.100	0.974	0	0.741

APPENDIX F: DESCRIPTORS IN THE AQUATIC TOXICITY MODELS

Table F1. Descriptors appeared in the algae model (Eq. 4.9)

Descriptor	Type	Meaning of descriptor
SPAM	Geometrical descriptors	Average span R
Mor31p	3D-MoRSE descriptors	Signal 31/weighted by polarizability
NdsCH	Atom-type E-state indices	Number of atoms of type dsCH
CATS2D_02_AP	CATS 2D	CATS2D Acceptor-Positive at lag 02
B05[C-S]	2D Atom Pairs	Presence/absence of C-S at topological distance 5
F03[C-N]	2D Atom Pairs	Frequency of C-N at topological distance 3
MLOGP2	Molecular properties	Squared Moriguchi octanol-water partition coefficient
Hardness	Quantum chemical (energy)	Half of the energy difference between the lowest unoccupied and highest occupied molecular orbitals

Table F2. Descriptors appeared in the RTL-W1 model equation (Eq. 4.10)

Descriptor	Type	Meaning of descriptor
nRCOOH	Functional group counts	the number of aliphatic carboxylic acids
E HOMO	Quantum chemical	the highest occupied molecular orbital energy

Table F3. Descriptors appeared in *Dugesia japonica* model (Eq. 4.11)

Descriptor	Type	Meaning of descriptor
$\log K_{ow}$		Logarithm of n-Octanol/Water Partition Coefficient
GATS7p	2D autocorrelations	Geary autocorrelation of lag 7 weighted by polarizability
SpMaxA_G/D	3D matrix-based descriptors	normalized leading eigenvalue from distance/distance matrix (folding degree index)
CATS2D_08_DL	CATS2D	CATS2D Donor-Lipophilic at lag 08
Mor31s	3D-MoRSE descriptors	signal 31 / weighted by I-state