Peer-reviewed version available at *Foods* **2019**, *8*, 628; doi:10.3390/foods8120628

Article

Prediction of Terpenoids Toxicity Based on Quantitative Structure-Activity Relationships Model

Rosa Perestrelo [1], Catarina Silva [1], Miguel X. Fernandes [2] and José S. Câmara [1,3,*]

[1] CQM, Centro de Química da Madeira, Universidade da Madeira, Campus da Penteada, 9020-105 Funchal, Portugal; rmp@staff.uma.pt, cgsluis@staff.uma.pt

[2] BioLab, Instituto Universitario de Bio-Orgánica "Antonio González" (IUBO-AG), Universidad de La Laguna, C/ Astrofísico Francisco Sánchez 2, 38200 La Laguna, Spain; mfernandes@ull.es

[3] Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira, Campus da Penteada, 9020-105 Funchal, Portugal. jsc@staff.uma.pt

* Correspondence: jsc@staff.uma.pt; Tel.: +351 291705112; (J.S.C.)

Abstract: Terpenoids, including monoterpenoids (C_{10}), norisoprenoids (C_{13}) and sesquiterpenoids (C_{15}), constitute a large group of plant-derived naturally occurring secondary metabolites which chemical structure is highly diverse. A quantitative structure-activity relationship (QSAR) model to predict the terpenoids toxicity and to evaluate the influences of their chemical structure, was developed in this study, by assessing the toxicity of 27 terpenoid standards using Gram-negative bioluminescent *Vibrio fischeri*, in real time. Under the test conditions, at concentration of 1 μM, the terpenoids showed a toxicity level lower than five %, with exception of geraniol, citral, (S)-citronellal, geranic acid, (±)-α-terpinyl acetate and geranyl acetone. Moreover, the standards tested displayed a toxicity level higher than 30 % at concentration of 50 to 100 μM, with the exception of (+)-valencene, eucalyptol, (+)-borneol, guaiazulene, β-caryophellene and linalool oxide. Regarding the functional group, the terpenoids toxicity was observed in the following order: alcohol > aldehyde ~ ketone > ester > hydrocarbons. CODESSA software was employed to develop the QSAR models based on the correlation of terpenoids toxicity and a pool of descriptors related to each chemical structure. The QSAR models, based on t-test values, showed that terpenoids toxicity was mainly attributed to geometric (e.g., asphericity) and electronic (e.g., max partial charge for a C atom [Zefirov's PC]) descriptors. Statistically, the most significant overall correlation was the four-parameter equation with training and test coefficient correlation higher than 0.810 and 0.535, respectively, and square coefficient of cross-validation (Q^2) higher than 0.689. According to the obtained data, the QSAR models are a suitable and a rapid tool to predict the terpenoids toxicity in a diversity of food products.

Keywords: terpenoids; *Vibrio fischeri*; toxicity; QSAR; heuristic method

1. Introduction

Monoterpenoids (C_{10}), norisoprenoids (C_{13}) and sesquiterpenoids (C_{15}) constitute a large group of plant-derived naturally occurring secondary metabolites with a highly diverse chemical structure, having various biological activities in addition to a wide range of applications including its use as agricultural products, flavorings, pharmaceuticals and fragrances [1]. From a health point of view, terpenoids are known for their antibacterial, anticancer, anti-inflammatory, anthelmintic, antiviral and antimalarial properties [2–5]. Terpenoids can function as antimicrobial agents to protect their natural hosts, with antibacterial activity occurring via disruption of the lipid membrane, resulting in alteration of membrane organization and functions [1,6]. As a result of lipophilic compounds partitioning into the lipid bilayer damage occurs in the cell membrane by impairing vital functions (e.g., ions, metabolites, lipids, and proteins loss; dissipation of the pH gradient and electrical potential) [6–8]. However, enzymes and DNA have also been mentioned as possible targets, as lipophilic compounds tend to associate with the hydrophobic core of several proteins leading to conformational

changes, and consequently protein inactivation [6]. The toxicity level depends on the interaction with membrane constituents, concentration and location. Relatively to the lipophilic compounds accumulation, which could occur at varying depths in the lipid bilayer, it depends on the compound hydrophobicity, as well as the influence of membrane composition on the incorporation of terpenoids into the system, or the effect of external factors (e.g., temperature on terpene penetration ability) [8].

Several *in vivo* assays are available to measure chemical toxicity. Nevertheless, these experimental assays are expensive, labor-intensive and time-consuming, which encourages the development of alternative more reliable, sensitive and quick bioassays [9]. In recent years, *Vibrio fischeri* (Gram negative bacterium) based on bioluminescence inhibition assay has been widely used to perform toxicity measurement, due to its good reproducibility, sensitivity, cost-effective, ease of operation and efficient ethical alternative to testing on higher species [10]. Researchers have reported the *V. fischeri* bioluminescence assay as the most sensitive across a wide range of chemicals compared to other bacterial assays, such as nitrification inhibition, respirometry, ATP luminescence, and enzyme inhibition [11,12]. This strain is also commercially available in several test kits i.e., Microtox, Aboatox, LUMIStox and ToxAlert [13].

Quantitative Structure Activity Relationship (QSAR) analysis is usually used to develop mathematical models that relate small variations of chemical structure, parameterized by empirical physicochemical or theoretical molecular descriptors, to biological activity [14]. Different types of numerical molecular descriptors have been employed, which are related to constitutional, topological, geometrical, electronic and quantum chemical origin [15]. Nevertheless, several steps should be taken into consideration to develop a robust and sensitive QSAR model, such as (*i*) understanding the interaction mechanism between chemical and biological system, (*ii*) selection relevant descriptors set that describe the relationship between the chemical and activity/property under consideration, and (*iii*) selection of the statistical tools [15–17].

Some studies have been performed on the relationship between toxicity and chemical structure of several compounds, and QSARs models were developed to predict *V. fischeri* toxicity for specific groups using molecular and physicochemical descriptors [18–23]. The toxicity of narcotic compounds against *V. fischeri* was also predicted using molecular connectivity indices (topological descriptors), and the data obtained suggested that the degree of branching and the compounds electronic characteristic have a dominant role in the toxicity level [20]. Topological, electronic and log P descriptors has also been used to predict the toxicity of organic pollutants against *V. fischeri* [19]. Charge distribution (e.g., max partial charge for a C atom [Zefirov's PC]) and geometric (e.g., shadow parameter) descriptors were used by Couling et al. [21] to assess the toxicity of a diversity of ionic liquids against *V. fischeri* and Daphnia magna. In addition, Das et al. [18] developing a predictive QSAR models for the ecotoxicity of ionic liquids using the bacteria *V. fischeri* as an indicator response species. The aim of the current study was to evaluate the toxicity of 27 terpenoids (16 mono-, 8 sesquiterpenic compounds, 3 norisoprenoids) at different concentrations (1, 10, 50 and 100 μM) and incubation times (0, 20, 40, 60, 80 and 100 min) using *V. fischeri* bioluminescence inhibition assay. The previously experimental data set obtained was then used to develop QSAR models using CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) software to predict the terpenoids related chemical structure toxicity.

2. Materials and Methods

2.1. Reagents

Ethanol (99.9 %), potassium dihydrogen phosphate (KH_2PO_4, 99 %), glycerol (87 %), peptone from casein, meat extract and tryptic soy agar (TSA) were supplied from Merck (Darmstadt, Germany), whereas agar was obtained from Liofilchem (Teramo, Italy). Anhydrous sodium carbonate (99.8 %), sodium chloride (NaCl, 99 %), sodium hydroxide (NaOH, 98 %) and potassium chloride (KCl, 99 %) were purchased from Panreac (Barcelona, Spain), whereas sodium dihydrogen phosphate dihydrate (Na_2HPO_4 $2H_2O$, 99 %) was supplied from Fluka (Buchs, Switzerland).

4.2. Terpenoids standards

Figure 1 shows the terpenoids chemical structure used as authentic standards to evaluate the toxicity. Nerol (90 %), β-caryophyllene (98.5 %), (-)-α-cedrene (99 %), (-)-α-neoclovene (95 %), (+)-valencene (70 %), (Z)-nerolidol (95 %), and (-)-α-bisabolol (95 %) were supplied from Fluka (Buchs, Switzerland). *p*-Cymene (99 %), (*R*)-(+)-limonene (97 %), (+)-borneol (97 %), eucalyptol (99 %), geraniol (98 %), linalool (98.5 %), α-terpeniol (95 %), β-citronellol (95 %), (-)-menthol (99 %), (*R*)-carvone (98 %), citral (95 %), (*S*)-citronellal (96 %), geranic acid (85 %), linalool oxide (97 %), (±)-α-terpinyl acetate (90 %), β-ionone (97 %), geranyl acetone (98 %), (±)-theaspirane (90 %) and (*E*,*E*)-farnesol (96 %) from Sigma–Aldrich Química S.A. (Madrid, Spain), whereas guaiazulene (98 %) was supplied from TCI Europe N.V. (Zwijndrecht, Belgium). For each terpenoid standard, an ethanolic stock solution was prepared (50 mM). From each stock solution, working solutions were prepared by diluting adequate amounts in order to obtain a final concentration of 0.2, 1, 10 and 20 mM. All solutions were stored at -20 ºC.

4.3. Assessment terpenoids toxicity

V. fischeri terpenoids exposures were conducted according to the methodology previously described [24]. The bacterial strains used in this work was a bioluminescent marine bacterium V. fischeri ATCC 49387 (USA), that produce light without the addition of exogenous substrates, and its light emission is directly proportional to their metabolic activity. Fresh plate cultures of bioluminescent *V. fischeri* were maintained in solid BOSS medium (1 % peptone, 0.3 % beef extract, 0.1 % glycerol, 3 % NaCl, 1.5 % agar, pH 7.3) at 4 ºC. A NaCl concentration range from 20 to 40 g/L is needed to maintain the osmotic pressure of cells required to natural light emission to occur. Before each bioassay, one isolated colony was asseptically inoculated in 30 mL of liquid BOSS medium, and grew during 16 h at 25 ºC under constant stirring (120 rpm). An aliquot of this culture (240 µL) was subcultured in 30 mL of BOSS medium, and grew overnight at 25 ºC under stirring (120 rpm) to reach an optical density (OD_{620}) of ≈ 1.0, corresponding to ≈ 10^8 CFU/mL. For bioassays purpose, an overnight culture of *V. fischeri* was used after a ten-fold dilution in phosphate buffered saline (PBS: 30 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 and 0.24 g KH_2PO_4 per litre; pH 7.4) to achieve a final concentration of 10^7 CFU/mL.

For each terpenoids experiment, 10 mL of bacterial suspension were aseptically distributed in 100 mL acid-washed and sterilized glass beakers and 50 µL of each standard working stock solution (0.2, 2, 10 and 20 mM in hydroalcoholic solution) was added in order to achieve a final concentration of 1, 10, 50 and 100 µM, respectively. Then, all beakers were wrapped with aluminium foil to protect from light exposure and incubated under 120 rpm stirring at 20-25ºC. A control experiment, consisting of bacterial suspension and hydroalcoholic solution, instead of terpenoids, was carried out simultaneously. Aliquots of 500 µL of the C_{10} and C_{15} terpenic compounds and norisoprenoids standard and of the control were collected at different times (0, 20, 40, 60, 80 and 100 min) and the bioluminescence signal (peak wavelength detected at 420 nm, standard range 300 - 650 nm) was measured on a luminometer (TD-20/20 Luminometer, Turner Designs, Inc., USA). Three independent assays were performed for each component and for the control and the results were averaged.

4.4. Calibration of bioluminescent signal and viable cell numbers

The correlation between the colony-forming units (CFU) and the bioluminescent signal (in relative light units, RLU) of *V. fischeri* was performed. For this purpose, eighth-fold serial dilutions of the culture were prepared in PBS with 3 % of NaCl. The non-diluted (100) and the diluted aliquots were spread plated in Triptic Soy Agar (TSA) with 3 % of NaCl (100 µL) to determine the number of viable cells (CFU/mL), and simultaneous were read on the luminometer (500 µL) to determine the

bioluminescence signal. Both experiments were performed in triplicate and the results were averaged.

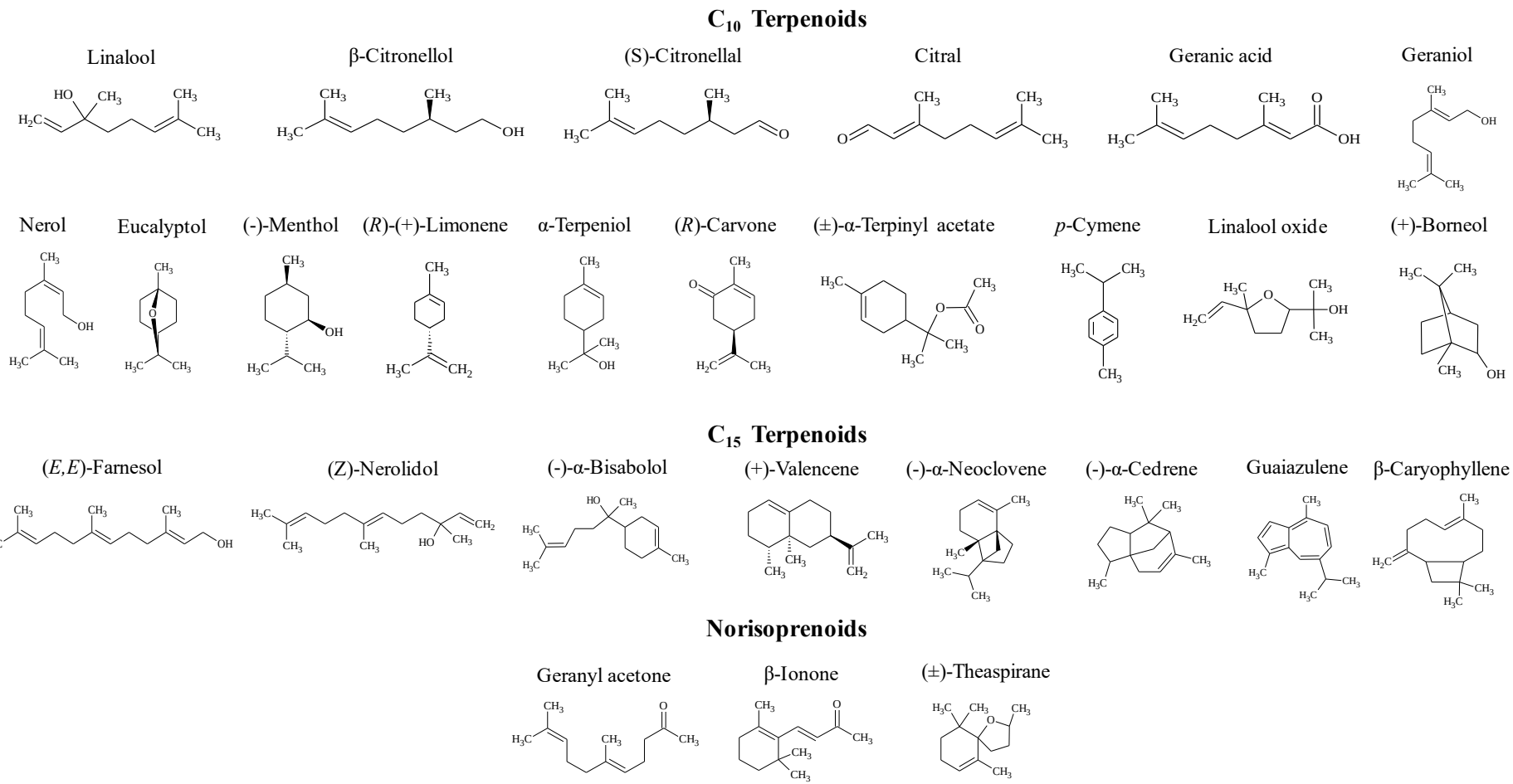

Figure 1. Terpenoids chemical structure.

4.5. QSAR model development

4.5.1. Geometry optimization and molecular descriptors calculation

The three-dimensional terpenoids chemical structures were drawn and pre-optimized using the AMBER force field model available in HYPERCHEM 7.0 software (Hypercube Inc, Gainsville, FL, USA). The final molecular geometries were refined using the quantum chemical program package MOPAC 6.0. The AM1 parameterization with eigenvector following geometry optimization procedure at a precision level 0.01 kcal/Å gradient norm was used to calculate electronic and thermodynamic descriptors.

The CODESSA (Semichem Inc, Shawnee, KS, USA) was used to calculate a pool of different molecular descriptors using MOPAC output files, HyperChem structure files and additional descriptor calculated using DRAGON software package [25]. In total, more than 280 molecular descriptors were generated for each structure, which could be organized into five groups, namely constitutional, topological, geometrical, electrostatic and quantum chemical. These molecular descriptors contain information about the connections, shape, symmetry, charge distribution and quantum chemical properties of the chemical structures under study.

4.5.2. QSAR models development and validation

An important step of the QSAR model development is the selection of the best multilinear regression equation among a given descriptor set. Once molecular descriptors are calculated, the selection was performed using heuristic method (HM) available in the framework of the CODESSA software, which reduces the descriptors pool by eliminating that: (*i*) were not available for all structure studied; (*ii*) have a constant value for all structure studied; (*iii*) the F-value was below 1; (*iv*) the t-test value lower than 0.1 at a probability level of 0.05; (*v*) highly correlated descriptors provide approximately identical information, if their pair-wise correlation coefficient exceeded 0.80 [26]. The selected descriptors were then used for developing the QSAR prediction models by multiple linear regressions (MLRs), with a training subset composed by 22 terpenoids. The predictive power of the resulting models was evaluated by a test subset of five terpenoids representative biological activity of data set. For the 22 terpenoids training subset, not more than four descriptors were considered for correlation analysis, thereby keeping the ratio to a maximum of 4:1 [27].

The QSAR models derived from MLR analysis were used for further validation study (in order to select the reliable and robust models) by taking into account of highest squared correlation coefficient (r^2), square coefficient of cross validation (Q^2), Fisher F-criterion value (ratio of regression and residual variances and reflects the significance of the model) and Student t-test (reflects the significance of the parameter within the model), as well as the lowest standard deviation (S). Generally, Q^2 is used as a criterion of both robustness and predictive ability of the QSAR model. Many researches considered high Q^2 (for instance, Q^2 higher than 0.50) as an indicator or even as the ultimate proof of the high predictive power of the QSAR model [25,28].

3. Results and discussion

3.1. Toxicity of terpenoids

Previously terpenoids toxicity assessment, a correlation between the colony-forming units (CFU) and the bioluminescent signal (in relative light units, RLU) of overnight cultures of *V. fischeri* bioluminescent strain was performed in order to evaluate the viable bacterial abundance. A linear correlation was observed (**Figure 2**), which reflect the viable bacterial abundance. This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn. **Figures 3** to **5** shows the inhibitory percentage of *V. fischeri* exposed to 27 terpenoids (16 mono-, 8 sesquiterpenic compounds and 3 norisoprenoids) at different concentrations (1, 10, 50 and 100 μM) and incubation times (0, 20, 40, 60, 80 and 100 min).

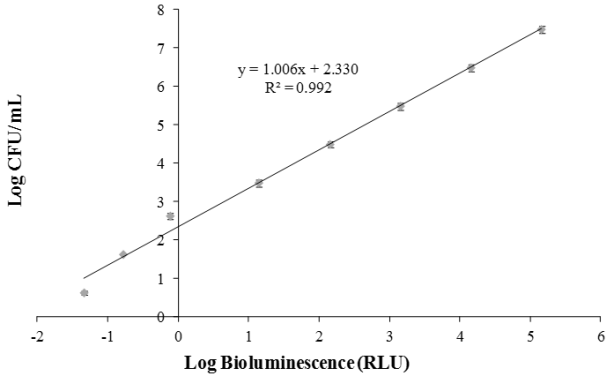

Figure 2. Relationship between the bioluminescence signal and viable counts of an overnight culture of *V. fischeri* (≈10^9 CFU/mL) serially diluted in PBS with 3 % of NaCl. Bioluminescence is expressed in relative light units (RLU) and viable counts in CFU/mL. Values represent the mean of two independent experiments; error bars indicate the standard deviation.

At concentration of 1 μM, geranic acid (20 %), (±)-α-terpinyl acetate (11 %), citral (9 %) and (*S*)-citronellal (8 %) showed higher toxicity level than the remaining terpenoids tested. At concentration of 10 μM, β-citronellol (52 %) showed a considerable toxicity level, followed by (±)-α-terpinyl acetate (32 %), β-ionone (31 %), geranyl acetone (28 %) ~ (Z)-nerolidol (28 %) ~ limonene (28 %), geranic acid (26 %) ~ (-)-α-bisabolol (26 %), (S)-citronellal (25 %), and the remaining terpenoids under study showed toxicity lower than 21 %. The results showed that toxicity was proportional to standard concentration, and at concentration of 100 μM, the majority of terpenoids tested showed toxicity higher than 50 %, with exception of (+)-valencene (14 %), eucalyptol (15 %), (+)-borneol (16 %), guaiazulene (16 %), β-caryophellene (19 %), linalool oxide (20 %), (-)-menthol (29 %), (+)-theaspirane (30 %), (R)-carvone (39 %) and (-)-α-neoclovene (46 %). Regarding to incubation time, in terms of toxicity, no remarkable differences were observed between 20 and 100 min, which means that greater toxicity occurs during the first 20 min. For this fact, this incubation time was selected to develop the QSAR models to predict the terpenoids toxicity. An overview could be achieved based on relationship of terpenoids toxicity and their chemical structure as well as functional groups. Concerning to the functional groups, the highest toxicity level was observed in the following order: alcohol (e.g., geraniol) > aldehyde (e.g., (S)-citronellal) ~ ketone (e.g., geranyl acetone) > ester (e.g., (±)-α-terpinyl acetate) > hydrocarbons (e.g., (+)-valencene). The toxicity of some C_{10} and C_{15} terpenoids alcohols has been previously reported [29], being the presence of hydroxyl groups crucial to toxicity level, suggesting that the binding sites may contain both hydrogen bond donors as well as hydrogen bond receptors [30]. According to the obtained data, the importance of hydroxyl group in terpenoids chemical structure was confirmed in terms of toxicity, when α-terpineol was compared to eucalyptol. The toxicity of terpenoids chemical structure could also increase due to the presence of oxygen related function (e.g., geranyl acetone, β-ionone, (*S*)-citronellal, citral). The presence of these functional groups increases the structure electronegativity, which may interfere in biological process involving electron transfer and react with vital nitrogen components, such as proteins and nucleic acid, consequently inhibit the bacterial growth [31]. The presence of the acetate moiety in the terpenoid chemical structure was also crucial to increase toxicity, which was confirmed when their activity of α-terpineol against *V. fischeri* was compared to (±)-α-terpinyl acetate (**Figure 3**). Similar results were reported to geraniol and (+)-borneol, where their toxicity was lower than the acetates against a diversity of bacteria [31]. The bacterial activity also depends on the alkyl substituent on the ring structure, which could be confirmed when (*R*)-(+)-limonene (alkenyl substituent) was compared to *p*-cymene (alkyl substituent). The presence of a double bound on C_{10}, C_{15} terpenoids and norisoprenoids chemical structure contributed to toxicity increases.

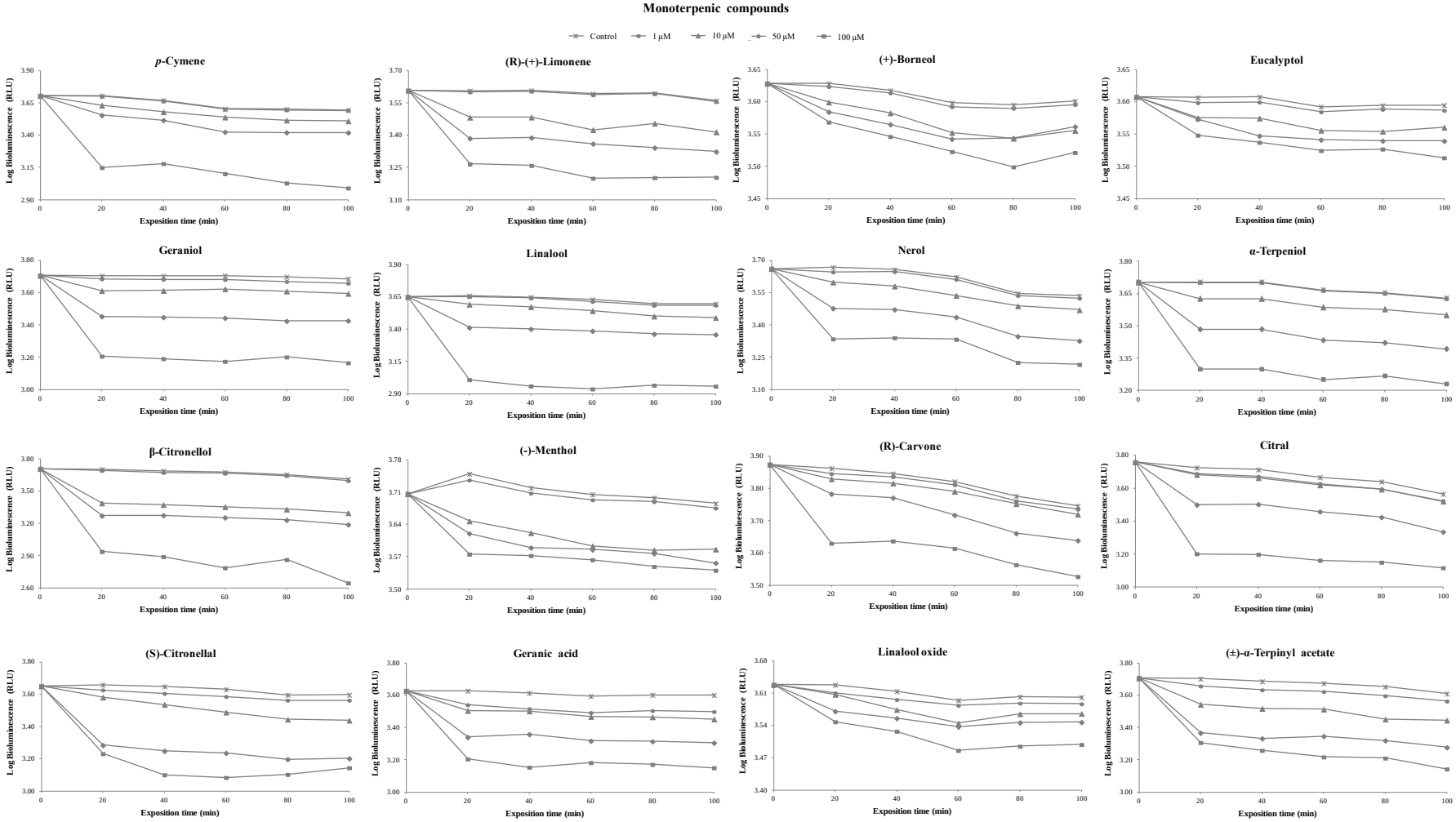

Figure 3. Bioluminescence monitoring of *V. fischeri* exposed to monoterpenic compounds at different concentration. The values are expressed as the means of three independent experiments.

Finally, the terpene hydrocarbons (e.g., (+)-valencene, guaiazulene, β-caryophyllene) compared to the other tested terpenoids in this study showed low toxicity, which could be explained by their low water solubility that limits their diffusion through the medium. This data are in agreement with previous studies, where C_{10} and C_{15} terpene hydrocarbons tend to be relatively inactive independently of their chemical structure, due to their limited hydrogen capacity and water solubility [32]. Their action site appeared to be at the lipid bilayer, caused by biochemical mechanism catalyzed by the lipid bilayers of the cell. These processes included the inhibition of electron transport, protein translocation, phosphorylation steps, and other enzyme-dependent reactions.

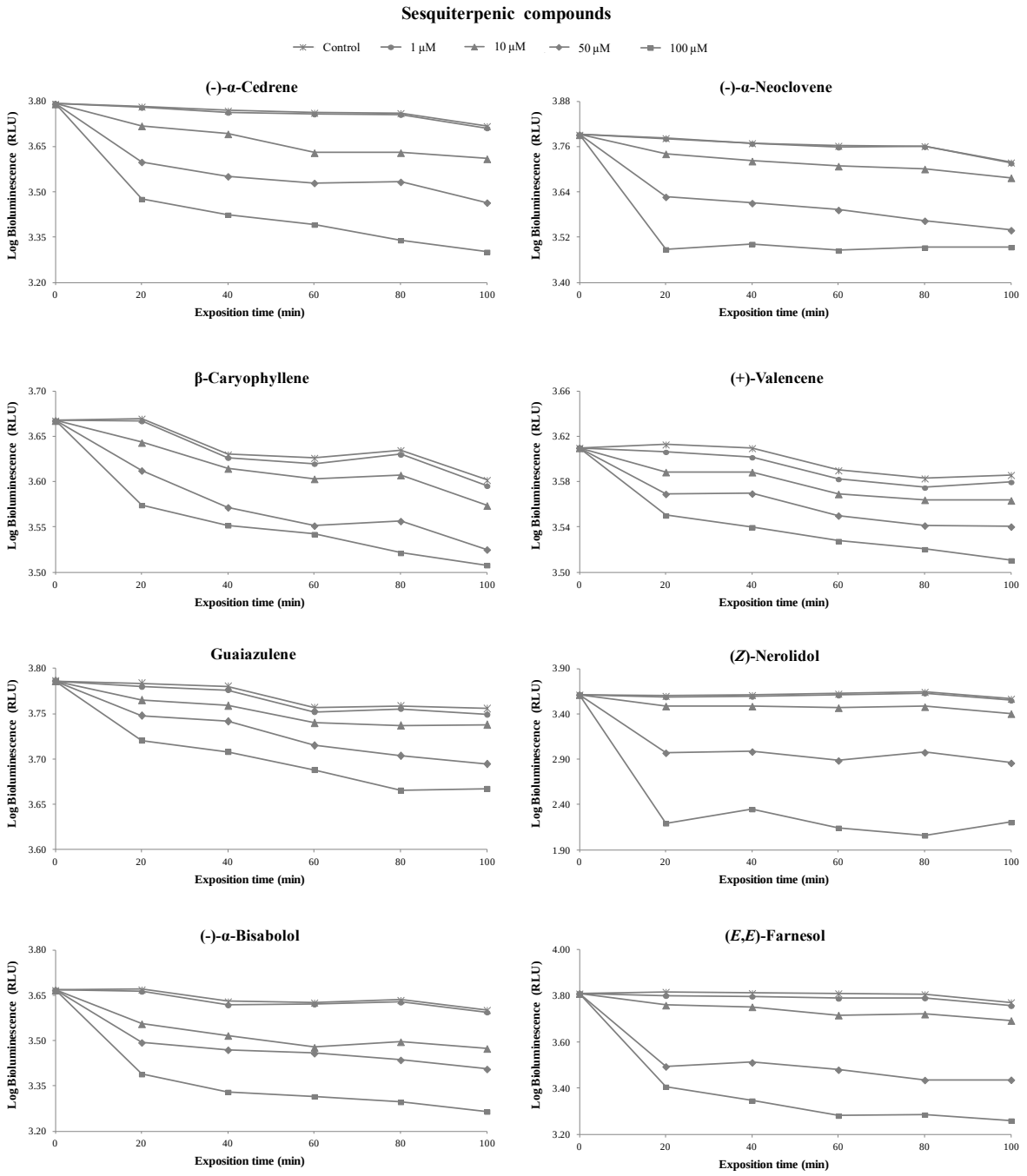

Figure 4. Bioluminescence monitoring of *V. fischeri* exposed to sesquiterpenic compounds at different concentration. The values are expressed as the means of three independent experiments.

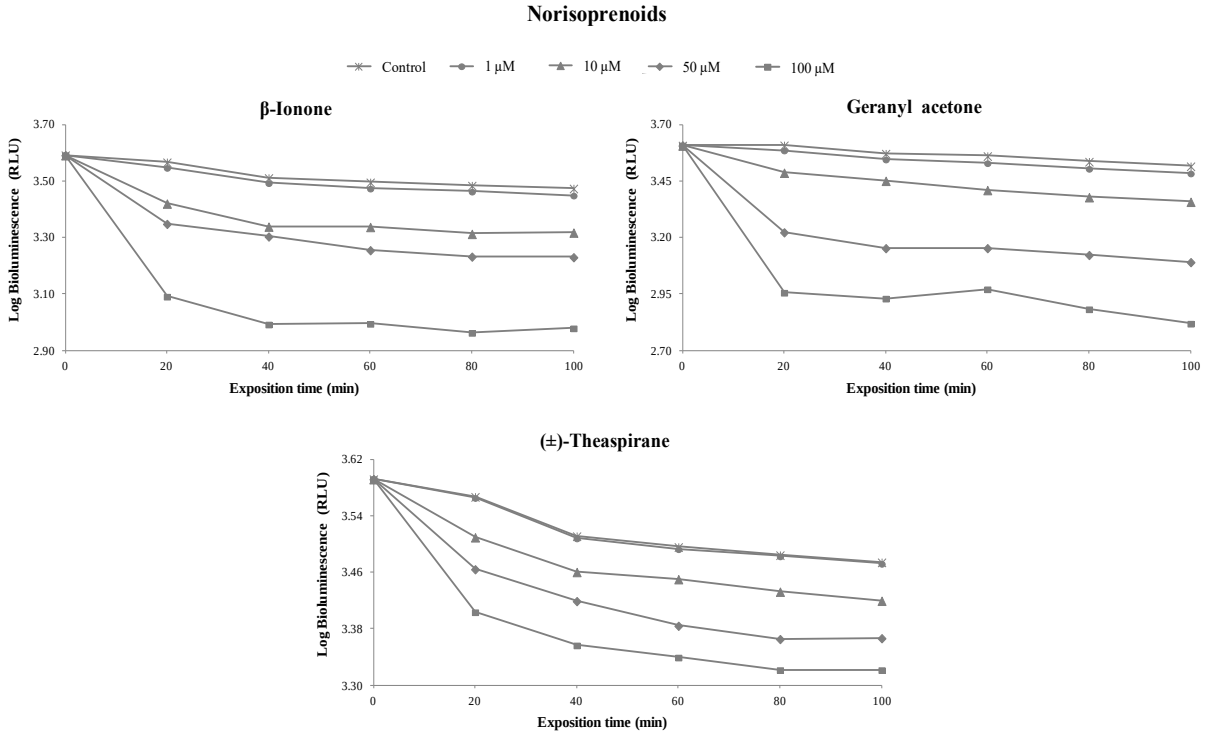

Figure 5. Bioluminescence monitoring of *V. fischeri* exposed to norisoprenoids at different concentration. The values are expressed as the means of three independent experiments.

The presence of these functional groups increases the structure electronegativity, which may interfere in biological process involving electron transfer and react with vital nitrogen components, such as proteins and nucleic acid, consequently inhibit the bacterial growth [31]. The presence of the acetate moiety in the terpenoid chemical structure was also crucial to increase toxicity, which was confirmed when their activity of α-terpineol against *V. fischeri* was compared to (±)-α-terpinyl acetate (**Figure 3**). Similar results were reported to geraniol and (+)-borneol, where their toxicity was lower than the acetates against a diversity of bacteria [31]. The bacterial activity also depends on the alkyl substituent on the ring structure, which could be confirmed when (*R*)-(+)-limonene (alkenyl substituent) was compared to *p*-cymene (alkyl substituent). The presence of a double bound on C_{10}, C_{15} terpenoids and norisoprenoids chemical structure contributed to toxicity increases. Finally, the terpene hydrocarbons (e.g., (+)-valencene, guaiazulene, β-caryophyllene) compared to the other tested terpenoids in this study showed low toxicity, which could be explained by their low water solubility that limits their diffusion through the medium. This data are in agreement with previous studies, where C_{10} and C_{15} terpene hydrocarbons tend to be relatively inactive independently of their chemical structure, due to their limited hydrogen capacity and water solubility [32]. Their action site appeared to be at the lipid bilayer, caused by biochemical mechanism catalyzed by the lipid bilayers of the cell. These processes included the inhibition of electron transport, protein translocation, phosphorylation steps, and other enzyme-dependent reactions.

3.2. QSAR models to predict the toxicity of terpenoids

The heuristic method (HM) was applied to generate QSAR models with four descriptors. A subset composed by 22 terpenoids was built, and the remaining five terpenoids were used as external validation subset, and the HM results are shown in **Table 1**. The QSAR models performed well, with a training correlation coefficient ($r^2_{training}$) and test (r^2_{test}) subset higher than 0.810 and 0.535, respectively, and square coefficient of cross validation (Q^2) values higher than 0.689, which suggests high predictive power. Although, the QSAR models developed to predict the terpenoids related chemical structure toxicity were characterized by good statistical parameters, such as $r^2_{training}$, r^2_{test} and Q^2, a good QSAR model fit depends on the experimental data quality. An extreme outlier was found

in the QSAR model generated for terpenoids concentration of 10 μM, and for this QSAR model (*Z*)-nerolidol (outlier) was removed in order to improve the statistical result (**Table 1**). The four descriptors involved in the QSAR models obtained for the each terpenoids concentration are listed in **Table 1**, and include constitutional, topological, geometrical, electrostatic and quantum chemical descriptors.

Table 1. QSAR models obtained for the different terpenoids concentration against *V. fischeri* bacterial at exposition time of 20 min.

[Terpenoids] (μM)	Nº	B	*t*-Test	Molecular descriptors	Statistical parameters
1	0	18.36	8.56	Intercept	$r^2_{training}$ = 0.952
	1	209.81	14.73	Max partial charge for a C atom [Zefirov's PC]	r^2_{test} = 0.923
	2	-10.28	-8.28	Max atomic orbital electronic population	F = 84.14
	3	0.93	6.70	Kier shape index (3rd order)	s^2 = 1.05
	4	-0.11	-3.90	WNSA-1 Weighted PNSA (PNSA1×TMSA/1000) [Zefirov's PC]	Q^2 = 0.900
10[a]	0	31.85	7.68	Intercept	$r^2_{training}$ = 0.873
	1	42.92	8.06	Asphericity	r^2_{test} = 0.6987
	2	-0.30	-6.50	PNSA-1 Partial negative surface area [Zefirov's PC]	F = 27.57
	3	148.82	4.50	Max partial charge for a C atom [Zefirov's PC]	s^2 = 11.25
	4	-3.92	-4.48	Log P	Q^2 = 0.794
50	0	39.71	1.07	Intercept	$r^2_{training}$ = 0.810
	1	110.40	7.40	Asphericity	r^2_{test} = 0.535
	2	7.53	3.71	Kier&Hall index (2nd order)	F = 18.17
	3	-0.28	-2.81	PNSA-1 Partial negative surface area [Zefirov's PC]	s^2 = 62.05
	4	-77.99	-1.95	Min atomic orbital electronic population	Q^2 = 0.689
100	0	19.48	0.94	Intercept	$r^2_{training}$ = 0.846
	1	195.98	8.58	Asphericity	r^2_{test} = 0.676
	2	21.01	0.21	Kier&Hall index (2nd order)	F = 23.39
	3	-1.17	-4.54	XY shadow	s^2 = 103.69
	4	-72.04	3.96	Relative number of single bonds	Q^2 = 0.734

[a] $N_{training}$ = 21, N_{test} = 5

For terpenoids concentration 1 μM (equation 1), the QSAR model was constituted by two electronic (max partial charge for a C atom [Zefirov's PC], Q^c_{max}, and WNSA-1 Weighted PNSA (PNSA1×TMSA/1000) [Zefirov's PC]), one topological (Kier shape index 3rd order, $^3\kappa$) and one quantum chemical (max atomic orbital electronic population, Max-OP) descriptors. According to the t-test, the most significant descriptor to predict terpenoids toxicity was Q^c_{max}, followed by max atomic orbital electronic population, $^3\kappa$ and WNSA-1. Q^c_{max} is an electronic descriptor calculated from Zerirov's electronegativity equation, and described the most positively charged C atom in the molecule that is usually connected to the electron withdrawing functional group or atom [33]. The Max-OP descriptor for a given atomic species in the molecule is a simplified index to describe the nucleophilicity of the molecule and could be interpreted as its ability to undergo oxidation and start degenerative metabolic process [34]. The negative coefficient obtained for this quantum chemical descriptor indicates that the terpenoids toxicity increased with the decrease of Max-OP magnitude.

In $^3\kappa$, the shape of molecule depends on the number of skeletal atoms, molecular branching, and the ratio of the atomic radius and the radius on the carbon atom in the sp^3 hybridization state [33]. The positive coefficient of $^3\kappa$ descriptor suggested that the increase in molecule branching and the presence of heteroatoms promoted the terpenoids toxicity. WNSA-1 descriptor describes the negative partial charge distribution information in the molecule and then could account for the electrostatic interaction between the compound and the receptor [33]. A negative coefficient of the WNSA-1 descriptor implies that the activity increases as the value of this descriptor decreases. As observed the electronic descriptors involved in this model indicated they are charged partial surface area

(CPSA) descriptors, which suggested that surface area alone as a geometric descriptor was not sufficient to predict terpenoids toxicity. This is in agreement with previous studies that use CPSA descriptors to assess the molecule lipophilicity [27].

For terpenoids concentration 10 μM (equation 2), QSAR model was constituted by one geometric (asphericity), two electronic (Q^{c}_{max}, PNSA-1 Partial negative surface area [Zefirov's PC]) and physicochemical (log P) descriptors. According to the t-test, these descriptors obey the following order of significance: asphericity > PNSA-1 > Q^{c}_{max} ~ log P. Asphericity (Ω) is a geometric descriptor which describes the molecule deviation from the spherical shape, and calculated from the eigenvalue λ_i of the inertia matrix [35]. The positive sign of asphericity indicates that terpenoids toxicity was promoted by linear (Ω = 1) and oblate (Ω ~ 1) molecules structure. PNSA-1 describes the sum of the surface area of negative atoms, and as observed for the first model (equation 1, **Table 1**), the negative sign of this electronic descriptor highlights that a decrease in the magnitude of PNSA favors the terpenoids toxicity. Again, Q^{c}_{max} showed a positively correlation with terpenoids toxicity. This result is in agreement with literature as the Gram-negative outer layer membrane is composed primarily by lipopolysaccharide molecules, and forms a hydrophilic permeability barrier providing protection against the effects of highly hydrophobic compounds [30].

At 50 μM terpenoids concentration (equation 3), QSAR model was also constituted by one geometric (asphericity), one topological (Kier&Hall index 2nd order, $^{2}\chi^{v}$), one electronic (PNSA-1) and one quantum chemical (min atomic orbital electronic population, Min-OP) descriptors. According to the t-test, these descriptors obey the following order of significance: asphericity > $^{2}\chi^{v}$ > PNSA-1 > Min-OP (equation 3, **Table 1**). As observed in QSAR model 2, the asphericity and PNSA-1 showed a positive and negative correlation, respectively, with terpenoids toxicity. $^{2}\chi^{v}$ is a valence connectivity topological descriptor, which reflects the branching molecule and also encodes the molecule size. The positive sign of this descriptor indicates that high molecular branching promote less London dispersion; consequently increases the terpenoids toxicity. The Min-OP descriptor for a given atomic species in the molecule is a simplified index to describe the electrophilic ability of the molecule and connected to the hydrogen donor capabilities of the molecule.

At 100 μM terpenoids concentration (equation 4, **Table 1**), based on t-test, the most significant descriptor in this model affecting the terpenoids toxicity was asphericity followed by XY shadow, number of single bonds (C_1) and $^{2}\chi^{v}$, which indicate that toxicity was effect by geometric, topological and constitutional descriptors, but not any electrostatic and quantum chemical descriptor. Again, asphericity showed a positive correlation with terpenoids toxicity, which indicates that toxicity was favored by the increase of asphericity magnitude, as observed in QSAR models 2 and 3 (**Table 1**). XY shadow is defined as the area of shadows of the molecule as projected on the XY plane by the orientation of the molecule in the space along the axes of inertia, which characterizes the size and geometrical shape of the molecule. Thus, it can act as a descriptor of van der Waals and dispersion interactions between chemical compound and lipid [36].

The QSAR models generated for each concentration suggest that the charge distribution over the molecule as well as shape, size and orientation of substituents remarkably influence the terpenoid toxicity. Moreover, it can be concluded that the developed models corresponding to the terpenoids concentration of 10, 50 and 100 μM, followed the same tendency, as according to the t-test values the toxicity was mainly affected by steric effects (e.g., asphericity), being β-citronellol (Ω = 0.71), (*E*,*E*)-farnesol (Ω = 0.69), (*S*)-citronellal (Ω = 0.68), geranic acid (Ω = 0.63), (*Z*)-nerolidol (Ω = 0.63), geranyl acetone (Ω = 0.61), citral (Ω = 0.61), (-)-α-bisabolol (Ω = 0.60), geraniol (Ω = 0.58), nerol (Ω = 0.57), linalool (Ω = 0.56) and (±)-*α*-terpinyl acetate (Ω = 0.56) the most toxicity. In sum, the presence of hydroxyl group as well as oxygen related function is crucial to terpenoids toxicity level, since the presence of these functional groups increases the structure electronegativity, which may interfere in biological process involving electron transfer and react with vital nitrogen components (e.g., proteins, nucleic acids).

4. Conclusions

The current study reports the toxicity of terpenoids against *V. fischeri* bacteria. Concerning to the functional groups, the terpenoids toxicity decreased in the order alcohol (*e.g.*, geraniol) > aldehyde (*e.g.*, (*S*)-citronellal) ~ ketone (*e.g.*, geranyl acetone) > ester (*e.g.*, (±)-α-terpinyl acetate) > hydrocarbons (e.g., (+)-valencene). The high sensibility of *V. fischeri* to the cytotoxic effect of terpene alcohols could be explained by the involvement of the hydroxyl group in the formation of hydrogen bonds with the membrane polar part. Whereas, the low sensibility of *V. fischeri* to the cytotoxic effect of hydrocarbons terpenes could be explained by the fact of Gram-negative outer layer membrane be composed primarily by lipopolysaccharide molecules, and forms a hydrophilic permeability barrier providing protection against the effects of highly hydrophobic compounds.

The previous experimental data set was used to generate the QSAR models. The models performed well, with high significant correlation obtained using heuristic method indicating that a combination of different molecular descriptors types originated the best correlation which could be used to predict the terpenoids related chemical structure toxicity. Among the obtained models common descriptors were found namely two electronic (max partial charge for a C atom [Zefirov's PC], PNSA-1 Partial negative surface area [Zefirov's PC]), one geometric (asphericity) and one topological (Kier&Hall index 2nd order). Their statistical significance depends on the terpenoids concentration, as it was observed for the lowest concentration (1 μM) tested the most significant is an electronic descriptor (max partial charge for a C atom [Zefirov's PC]), whereas for the remaining tested concentrations is a geometric (asphericity) descriptor. Both descriptors showed a positive correlation with toxicity, suggests that molecule branching, heteroatoms presence and electronegativity play dominant role in terpenoids toxicity, and the most potential terpenoids toxicity were β-citronellol, (*E*,*E*)-farnesol, (*S*)-citronellal, geranic acid, (*Z*)-nerolidol, (-)-β-bisabolol, geraniol, nerol, linalool, geranyl acetone, β-ionone, citral, geranic acid and (±)-α-terpinyl acetate. The developed QSAR models provided a suitable and rapid tool to predict the terpenoids toxicity present in a diversity of food products.

Author Contributions: Rosa Perestrelo and Catarina Silva did the experimental assays, performed manuscript writing and preparation; Miguel X. Fernandes develop the QSAR model, correction and editing of manuscript; José S. Câmara performed the conception of study, manuscript preparation, correction and editing of manuscript.

Funding: This research was funded by FCT-Fundação para a Ciência e a Tecnologia (projects PEst-OE/QUI/UI0674/2019, CQM, Portuguese Government funds), and Madeira 14-20 Program, project PROEQUIPRAM - Reforço do Investimento em Equipamentos e Infraestruturas Científicas na RAM (M1420-01-0145-FEDER-000008) and ARDITI-Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação, through the project M1420-01-0145-FEDER-000005 - Centro de Química da Madeira - CQM^{+} (Madeira 14-20).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vermaas, J. V; Bentley, G. J.; Beckham, G. T.; Crowley, M. F. Membrane permeability of terpenoids explored with molecular simulation. J. Phys. Chem. B **2018**, 122, 10349–10361.
2. Ludwiczuk, A.; Skalicka-Woźniak, K.; Georgiev, M. I. Terpenoids. In Pharmacognosy; Academic Press, **2017**; pp. 233–266 ISBN 9780128021040.
3. Mbaveng, A. T.; Hamm, R. Harmful and protective effects of terpenoids from african medicinal plants. In Toxicological Survey of African Medicinal Plants; Elsevier, **2014**; pp. 557–576 ISBN 9780128000182.
4. Shakya, A. K. Medicinal plants: Future source of new drugs. Int. J. Herb. Med. **2016**, 4, 59–64.
5. Nevzorova, Y. A.; Grossmann, J.; Trautwein, C. Anti-tumorigenic and anti-angiogenic effects of natural conifer Abies sibirica terpenoids in vivo and in vitro. Biomed. Pharmacother. **2017**, 89, 386–395.
6. Lim, C.; Park, S.; Park, J.; Ko, J.; Lee, D. W.; Hwang, D. S. Probing nanomechanical interaction at the interface between biological membrane and potentially toxic chemical. J. Hazard. Mater. **2018**, 353, 271–279.
7. Blaskó, Á.; Gazdag, Z.; Gróf, P.; Máté, G.; Sárosi, S.; Krisch, J.; Vágvölgyi, C.; Makszin, L.; Pesti, M. Effects of clary sage oil and its main components, linalool and linalyl acetate, on the plasma membrane of Candida albicans: an *in vivo* EPR study. Apoptosis **2017**, 22, 175–187.
8. Hąc-Wydro, K.; Flasiński, M. F.; Broniatowski, M.; Sołtys, M. Studies on the behavior of eucalyptol and terpinen-4-ol - natural food additives and ecological pesticides- in model lipid membranes. Langmuir **2017**, 33, 6916–6924.
9. Mobed, A.; Hasanzadeh, M.; Agazadeh, M.; Mokhtarzadeh, A.; Rezaee, M. A.; Sadeghi, J. Bioassays: The best alternative for conventional methods in detection of *Legionella pneumophila*. Int. J. Biol. Macromol. **2019**, 121, 1295–1307.
10. Abbas, M.; Adil, M.; Ehtisham-ul-Haque, S.; Munir, B.; Yameen, M.; Ghaffar, A.; Shar, G. A.; Asif Tahir, M.; Iqbal, M. *Vibrio fischeri* bioluminescence inhibition assay for ecotoxicity assessment: A review. Sci. Total Environ. **2018**, 626, 1295–1309.
11. Jarque, S.; Masner, P.; Klánová, J.; Prokeš, R.; Bláha, L. Bioluminescent *Vibrio fischeri* assays in the assessment of seasonal and spatial patterns in toxicity of xontaminated river sediments. Front. Microbiol. **2016**, 7, 1738.
12. Parvez, S.; Venkataraman, C.; Mukherji, S. A review on advantages of implementing luminescence inhibition test (*Vibrio fischeri*) for acute toxicity prediction of chemicals. Environ. Int. **2006**, 32, 265–268.
13. Kusumahastuti, D. K. A.; Sihtmäe, M.; Kapitanov, I. V.; Karpichev, Y.; Gathergood, N.; Kahru, A. Toxicity profiling of 24 l-phenylalanine derived ionic liquids based on pyridinium, imidazolium and cholinium cations and varying alkyl chains using rapid screening *Vibrio fischeri* bioassay. Ecotoxicol. Environ. Saf. **2019**, 172, 556–565.
14. Luis, P.; Garea, A.; Irabien, A. Quantitative structure–activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC_{50} (*Vibrio fischeri*). J. Mol. Liq. **2010**, 152, 28–33.
15. Pandith, A. H.; Giri, S.; Chattaraj, P. K. A comparative study of two quantum chemical descriptors in predicting toxicity of aliphatic compounds towards Tetrahymena pyriformis. Org. Chem. Int. **2010**, 2010, 1–17.
16. Shahlaei, M. Descriptor selection methods in quantitative structure–activity relationship studies: A review study. Chem. Rev. **2013**, 113, 8093–8103.
17. Moorthy, N. S. H. N.; Ramos, M. J.; Fernandes, P. A. Topological, hydrophobicity, and other descriptors on α-glucosidase inhibition: a QSAR study on xanthone derivatives. J. Enzyme Inhib. Med. Chem. **2011**, 26, 755–766.
18. Das, R. N.; Sintra, T. E.; Coutinho, J. A. P.; Ventura, S. P. M.; Roy, K.; Popelier, P. L. A. Development of predictive QSAR models for *Vibrio fischeri* toxicity of ionic liquids and their true external and experimental validation tests. Toxicol. Res. (Camb). **2016**, 5, 1388–1399.
19. Parvez, S.; Venkataraman, C.; Mukherji, S. Toxicity assessment of organic pollutants: Reliability of bioluminescence inhibition assay and univariate QSAR models using freshly prepared *Vibrio fischeri*. Toxicol. Vitr. **2008**, 22, 1806–1813.
20. 20. Agrawal, V. K.; Khadikar, P. V. QSAR study on narcotic mechanism of action and toxicity: a molecular connectivity approach to *Vibrio fischeri* toxicity testing. Bioorg. Med. Chem. **2002**, 10, 3517–3522.
21. Couling, D. J.; Bernot, R. J.; Docherty, K. M.; Dixon, J. K.; Maginn, E. J. Assessing the factors responsible for ionic liquid toxicity to aquatic organisms via quantitative structure–property relationship modeling. Green Chem. **2006**, 8, 82–90.

22. Ghanem, O. Ben; Mutalib, M. I. A.; Lévêque, J.-M.; El-Harbawi, M. Development of QSAR model to predict the ecotoxicity of *Vibrio fischeri* using COSMO-RS descriptors. Chemosphere **2017**, 170, 242–250.
23. Peric, B.; Sierra, J.; Martí, E.; Cruañas, R.; Garau, M. A. Quantitative structure–activity relationship (QSAR) prediction of (eco)toxicity of short aliphatic protic ionic liquids. Ecotoxicol. Environ. Saf. **2015**, 115, 257–262.
24. Alves, E.; Rodrigues, J. M. M.; Faustino, M. A. F.; Neves, M. G. P. M. S.; Cavaleiro, J. A. S.; Lin, Z.; Cunha, Â.; Nadais, M. H.; Tomé, J. P. C.; Almeida, A. A new insight on nanomagnet–porphyrin hybrids for photodynamic inactivation of microorganisms. Dye. Pigment. **2014**, 110, 80–88.
25. Beheshti, A.; Pourbasheer, E.; Nekoei, M.; Vahdani, S. QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm–multiple linear regressions. J. Saudi Chem. Soc. **2016**, 20, 282–290.
26. Lather, V.; Kairys, V.; Fernandes, M. X. Quantitative Structure-Activity Relationship models with receptor-dependent descriptors for predicting peroxisome proliferator-activated receptor activities of thiazolidinedione and oxazolidinedione derivatives. Chem. Biol. Drug Des. **2009**, 73, 428–441.
27. Grover, M.; Gulati, M.; Singh, B.; Singh, S. RP-HPLC determination of lipophilicity of 22 penicillins, their correlation with reported values and establishment of quantitative structure-log Kw relationships. QSAR Comb. Sci. **2005**, 24, 639–648.
28. Larsson, M.; van den Berg, M.; Brenerová, P.; van Duursen, M. B. M.; van Ede, K. I.; Lohr, C.; Luecke-Johansson, S.; Machala, M.; Neser, S.; Pěnčíková, K.; Poellinger, L.; Schrenk, D.; Strapáčová, S.; Vondráček, J.; Andersson, P. L. Consensus toxicity factors for polychlorinated dibenzo-p-dioxins, dibenzofurans, and biphenyls combining in silico models and extensive in Vitro screening of AhR-mediated effects in human and rodent cells. Chem. Res. Toxicol. **2015**, 28, 641–650.
29. Trombetta, D.; Castelli, F.; Sarpietro, M. G.; Venuti, V.; Cristani, M.; Daniele, C.; Saija, A.; Mazzanti, G.; Bisignano, G. Mechanisms of antibacterial action of three monoterpenes. Antimicrob. Agents Chemother. **2005**, 49, 2474–8.
30. Cristani, M.; D'Arrigo, M.; Mandalari, G.; Castelli, F.; Sarpietro, M. G.; Micieli, D.; Venuti, V.; Bisignano, G.; Saija, A.; Trombetta, D. Interaction of four monoterpenes contained in essential oils with model membranes: implications for their antibacterial activity. J. Agric. Food Chem. **2007**, 55, 6300–6308.
31. Group, M. P.; Ayrshire, S.; Estadual, U.; Julio, P.; Filho, D. M.; Safety, F. Antimicrobial agents from plants: antibacterial activity of plant volatile oils. J. Appl. Microbiol. **2006**, 88, 25–27.
32. Dambolena, J. S.; López, A. G.; Cánepa, M. C.; Theumer, M. G.; Zygadlo, J. A.; Rubinstein, H. R. Inhibitory effect of cyclic terpenes (limonene, menthol, menthone and thymol) on *Fusarium verticillioides* MRC 826 growth and fumonisin B1 biosynthesis. Toxicon **2008**, 51, 37–44.
33. Maran, U.; Sild, S. QSAR modeling of mutagenicity on non-congeneric sets of organic compounds. In Artificial Intelligence Methods And Tools For Systems Biology; Springer Netherlands: Dordrecht, **2004**; pp. 19–35.
34. Bruzzone, S.; Chiappe, C.; Focardi, S. E.; Pretti, C.; Renzi, M. Theoretical descriptor for the correlation of aquatic toxicity of ionic liquids by quantitative structure–toxicity relationships. Chem. Eng. J. **2011**, 175, 17–23.
35. Goudarzi, N.; Fatemi, M. H.; Samadi-Maybodi, A. Quantitative structure–properties relationship study of the ^{29}Si-NMR chemical shifts of some silicate species. Spectrosc. Lett. **2009**, 42, 186–193.
36. Liu, H. X.; Yao, X. J.; Zhang, R. S.; Liu, M. C.; Hu, Z. D.; Fan, B. T. Prediction of the tissue/blood partition coefficients of organic compounds based on the molecular structure using least-squares support vector machines. J. Comput. Aided. Mol. Des. **2005**, 19, 499–508.