Science of the Total Environment 610–611 (2018) 937–943

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

The quantitative structure–insecticidal activity relationships from plant derived compounds against chikungunya and zika *Aedes aegypti* (Diptera:Culicidae) vector

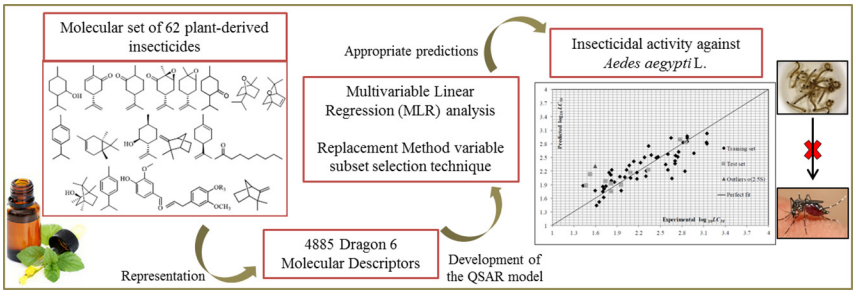
Laura M. Saavedra [a,*], Gustavo P. Romanelli [b,c], Ciro E. Rozo [d], Pablo R. Duchowicz [a,*]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, UNLP, Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina
[b] Centro de Investigación y Desarrollo en Ciencias Aplicadas "Dr. J.J. Ronco" (CINDECA), Departamento de Química, Facultad de Ciencias Exactas, CONICET, UNLP, Calle 47 No. 257, B1900AJK La Plata, Argentina
[c] Cátedra de Química Orgánica, Centro de Investigación en Sanidad Vegetal (CISaV), Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata, Calles 60 y 119 s/n, B1904AAN La Plata, Argentina
[d] Grupo de Investigaciones Ambientales para el Desarrollo Sostenible (GIADS), Universidad Santo Tomas, Seccional Bucaramanga, Carrera 18 No. 9-27. 680011 Bucaramanga, Colombia

HIGHLIGHTS

- Prediction of the insecticidal activity for sixty-two plant derived compounds against chikungunya and zika *A. aegypti* vector.
- QSAR models are suggested for modelling the acute toxicity of bioactive molecules using 4885 Dragon 6 molecular descriptors.
- The Replacement Method based on Multivariable Linear Regression is a reliable feature selection method in QSAR.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form 8 August 2017
Accepted 12 August 2017
Available online xxxx

Keywords:
Insecticidal activity
Chikungunya
Zika
Aedes aegypti
QSAR theory
MLR analysis

ABSTRACT

The insecticidal activity of a series of 62 plant derived molecules against the chikungunya, dengue and zika vector, the *Aedes aegypti* (Diptera:Culicidae) mosquito, is subjected to a Quantitative Structure–Activity Relationships (QSAR) analysis. The Replacement Method (RM) variable subset selection technique based on Multivariable Linear Regression (MLR) proves to be successful for exploring 4885 molecular descriptors calculated with Dragon 6. The predictive capability of the obtained models is confirmed through an external test set of compounds, Leave-One-Out (LOO) cross-validation and Y-Randomization. The present study constitutes a first necessary computational step for designing less toxic insecticides.

1. Introduction

The transmission of endemic diseases is mainly allocated to mosquitoes. There are reported alarming values of morbidity and mortality in tropical and subtropical regions,(Katritzky et al., 2008) where it is estimated that approximately 2.5 billion people live under the threat of

* Corresponding authors.
E-mail addresses: laurasaa0913@gmail.com (L.M. Saavedra), pabloducho@gmail.com (P.R. Duchowicz).

http://dx.doi.org/10.1016/j.scitotenv.2017.08.119
0048-9697/

various arbovirus types such as yellow fever, dengue and chikungunya fever.(Ocampoa et al., 2011) This last one has an important impact on public health in Brazil, Mexico, Colombia and Argentina, with 40,000 case reports of viral infection from 2013 to 2015.(World Health Organization (WHO), n.d.-a)

During the last years, mosquitoes have been responsible for the transmission of the zika virus (ZIKV) in Brazil and Colombia with 146,675 recognized cases.(World Health Organization (WHO), n.d.-b) The main infection reason is the proliferation of vectors such as *Aedes aegypti*, *Aedes leucocaelenus*, *Aedes albopictus* and *Aedes sabethes*. Owing to continuous climatic changes, these insects have increased their population, thus spreading to new territories where they have not been previously found.(Gillij et al., n.d.) Moreover, the effectiveness of pesticides used for diseases vector control has been increasingly affected by environmental conditions.(WHO, 2006; WHO, 2009)

The vectors propagation control method, proposed by the Pan American Health Organization (PAHO) during the fifties, has applied the organochlorine insecticide dichloro-diphenyl-trichloroethane (DDT) in twenty-one countries. Unfortunately, some years later, the vector population has become highly resistant to this pesticide, thus leading to its use ban. According to the National Pesticide Information Center (NPIC), DDT currently persists in the environment resulting in high toxicity to humans and animals as birds, fish and rats.(National Pesticide Information Center- NPIC; Oregon State University, 1999) In the eighties, Colombia has proposed the implementation of organophosphorus compounds (OPs) such as temephos, with a high insecticidal activity against *A. aegypti* in larval stage and also in others non-targeted animals.

Synthetic repellents have been developed for use as personal protection from bloodsucking insects. These substances have local action modes acting on the central nervous system (CNS) of the insect, causing deterrent effects that result in host avoidance. The widely used *N*,*N*-diethyl-toluamide (DEET) repellent,(Environment Protection Agency-EPA.; U. E. P. A., 1980) a compound of topical application, has some problems with his efficacy, limited protection time, irritation cases are reported, allergies and systemic intoxication in humans, and also resistance in arthropods, such as *Drosophila melanogaster* house fly are reported.(Bhattacharjee et al., 2005)

The need to find new environmentally friendly compounds having insecticidal properties leads to the use of natural products, particularly the essential oils (EOs) and plant extracts, which are complex mixtures of bioactive compounds possessing various biological properties.(Song et al., 2013; Ceferina et al., 2006) Secondary metabolites, such as terpenes, alkaloids and phenylpropanoids are abundant compounds in nature, present in fruits, leaves and flowers. They are involved in a wide range of applications and are found in cosmetics, therapeutic drugs and food additives. Furthermore, they have outstanding biological and organoleptic properties, such as *d*-limonene with inhibitory and insecticidal effects, pulgone with larvicidal property, menthol or linalool like effective insect repellents, and eugenol with antifungal activity.(Rice & Coats, 1994; Zhou et al., 2012; Carrasco et al., 2012)

The search for natural compounds with insecticidal activity against *A. aegypti* requires time, large budget and reagents for biological and clinical assays. In this sense, the possibility of employing a simple theoretical methodology for predicting biological, organoleptic or physicochemical properties of the compounds from knowing of its molecular structure represents a plausible tool in the rational design of novel bioactive molecules.

The Quantitative Structure–Activity Relationships (QSAR) theory is pioneer in the prediction of physicochemical properties or biological activities.(Hansch & Leo, 1995) Linear or non-linear mathematical models are established that include molecular descriptors characterizing relevant structural aspects of the compounds.(Katritzky et al., 1995) In fact, it has been reported that the assignation of the physicochemical meaning of the molecular descriptors could be assessed by considering the chemical orthogonal space of chemical reactivity descriptors,(Putz et al., 2017; Putz & Dudas, 2013) while the interaction mechanism, which may not be clear from the combined influence of numerical descriptors in the linear correlation, may be also pursued through a computational-conceptual algorithm for better understanding the chemical-biological interaction.(Putz & Dudas, 2013) QSAR studies are able to reduce time and costs in experimental measurements.(Ibezim et al., 2012)

In the present QSAR study, we predict the insecticidal activity against *A. aegypti* from plant derived molecules with known experimental data (62 molecules). The larvicidal activity is expressed as the median lethal concentration (LC_{50}), a standard measure of the toxicity of compounds, which measures the concentration at which 50% of third-instar larvae show lethal effect after in a specified period of the testing solutions. We apply the Replacement Method (RM) variable subset selection approach applied in the linear regression analysis of 4885 Dragon descriptors.(Duchowicz et al., 2006) In the last years, the RM technique has been successful for selecting relevant structural information and for establishing linear QSAR models with high predictive capability.(Duchowicz et al., 2008)

2. Materials and methods

2.1. Experimental data

The experimental LC_{50} insecticidal activities of 62 natural or semi-synthetic compounds are collected from the literature.(Santos et al., 2010; Santos et al., 2011; Scotti et al., 2014; Barbosa et al., 2012) For modelling purposes, such values is converted into logarithmic scale ($LC_{50} = \log LC_{50}$). Fig. 1 displays the heterogeneous molecular structures analyzed, involving terpenes, phenylpropanoids, ketones and oxygenated compounds. The complete list of LC_{50} values of the molecular set studied here is included in Table 1S of the Supplementary material.

2.2. Calculation of molecular descriptors

The initial conformations of the compounds are drawn in HyperChem for Windows.(HyperChem 7, n.d.) The structures are pre-optimized with the Molecular Mechanics Force Field (MM+), followed by the PM3 (Parametric Method-3) semi-empirical method to refine the structures using the Polak-Ribiere algorithm and a gradient norm limit of 0.01 kcal mol^{-1} A^{-1}.

Afterward, the molecular descriptors are calculated with the Dragon 6 software.(Talete srl, n.d.) The descriptors set contains 4885 variables and includes several types characterizing the multidimensional aspects of the chemical structure: constitutional, topological, geometrical, charge, GETAWAY (geometry, topology and atoms-weighted assembly), WHIM (weighted holistic invariant molecular descriptors), 3D-MoRSE (3D molecular representation of structure based on electron diffraction), walk and path counts, 2D and 3D autocorrelations, connectivity indices, burden eigenvalues, ETA indices, edge adjacency indices, radial distribution function, Randic molecular profiles, functional groups counts and atom-centred fragments. For the descriptor set, we exclude descriptors with constant or near-constant values, and those with at least one missing value. With this process a set containing 1738 linearly-independent descriptors is achieved.

2.3. Molecular descriptor selection in MLR

The Multivariable Linear Regression (MLR) technique has proven to be of multidisciplinary use and valuable applicability for establishing predictive QSAR models.(Duchowicz et al., 2008) Linear models are general and clearly show the effect of including/excluding descriptors in the equation, therefore, it is possible to suggest cause/effect relationships through such simple parallelisms. The main advantage of developing linear regression models is that they pose fewer over-fitting (over-training) problems, because the MLR method does not require too