Available online at www.sciencedirect.com

ScienceDirect

European Journal of Medicinal Chemistry 44 (2009) 260–273

EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY

http://www.elsevier.com/locate/ejmech

Original article

Larvicidal activities of some organotin compounds on mosquito larvae: A QSAR study

Corwin Hansch, Rajeshwar P. Verma*

Department of Chemistry, Pomona College, 645 North College Avenue, Claremont, CA 91711, USA

Received 17 October 2007; received in revised form 9 February 2008; accepted 12 February 2008

Abstract

Mosquitoes are not only the cause of nuisance by their bites but also transmit deadly diseases like malaria, filariasis, yellow fever, dengue, and Japanese encephalitis. In this paper, nine QSAR models were developed using different series of organotins with respect to their larvicidal activities against *Aedes aegypti* and *Anopheles stephensi* mosquito larvae. Internal [cross-validation (LOO-q^2), quality factor (Q), Fischer statistics (F), and Y-randomization] and external validation tests have validated all these QSAR models. QSAR results suggest that the two most important determinants for the toxicity are the hydrophobic (π) and Hammett electronic (σ^+) parameters of the substituents, and the kill mechanism is different for these two species of mosquito larvae. On the basis of QSAR (6), nine compounds **4a**–**4i** are suggested as potential synthetic targets.

Keywords: Hydrophobicity; Larvae; Mosquito; Molar refractivity; Organotins; QSAR

1. Introduction

Mosquitoes (one of the most important single group of insects) are not only the cause of nuisance by their bites but also transmit deadly diseases like malaria, filariasis, yellow fever, dengue, and Japanese encephalitis, which contribute significantly to poverty and social debility in tropical countries, causing millions of death every year [1–3]. Mosquitoes in the genus *Anopheles* are vectors of human malaria, and certain species of *Aedes* mosquito are responsible for the transmission of yellow fever, dengue and other pathogenic viruses [4]. Malaria is one of the most widespread infectious diseases in the world. This disease is estimated to kill between 1.5 and 2.7 million people every year, most of them African children under 5 years of age [5].

Repeated use of synthetic insecticides for mosquito control has fostered several environmental and human health concerns, including disruption of natural–biological control systems, resurgences in mosquito populations, widespread development of resistance, and undesirable effects on non-target organisms [2,3]. These problems have highlighted the need of new strategies for mosquito control. Plants may be an alternative source of insecticidal agents because they constitute a rich source of bioactive chemicals [6]. Much effort has been focused on plant extracts or phytochemicals as potential sources of commercial mosquito-control agents or bioactive chemical compounds [7–10]. However, insecticides of plant origin have been extensively used on agricultural pests and to a very limited extent, against insect vectors of public health importance. Thus, the current emphasis for chemists is on the production of novel types of insecticide that prevent insect resistance and that are environmentally friendly with relatively long-term effects. The development of these kinds of insecticides would be of worldwide interest.

It is well documented that the organotin compounds display strong biocidal activities. Most of these compounds are generally very toxic, even at low concentration. The toxic activities of organotin compounds are found to be dependent on the number and nature of the organic groups bound to the central Sn atom. It has been observed that compounds with three Sn–C bonds (R_3SnX) show the highest biological activity [11,12].

---

* Corresponding author. Fax: +1 909 607 7726.
E-mail address: rverma@pomona.edu (R.P. Verma).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.02.040

Recently, in the interest of developing a more effective insecticide against the *Aedes aegypti* (*Ae. aegypti*) and *Anopheles stephensi* (*An. stephensi*) mosquitos, several classes of organotins have been synthesized and screened against their larvae and the adult mosquitos. The toxicity studies indicated that the organotins are effective against these two species of mosquito larvae [13–17]. In this paper, nine QSAR (Quantitative structure–activity relationship) models were developed on four different series of organotin compounds (R_3SnX) with respect to their larvicidal activities against the second instar stage of the *Ae. aegypti* and *An. stephensi* mosquito larvae to understand their chemical–biological interaction.

In the past 44 years, the use of QSAR (one of the well developed areas in computational chemistry), since the advent of this methodology [18], has become increasingly helpful in understanding many aspects of chemical–biological interactions in drug and pesticide research, as well as in the areas of toxicology. This method is useful in elucidating the mechanisms of chemical–biological interaction in various biomolecules, particularly enzymes, membranes, organelles, and cells, as well as in humans [19–21]. It has also been utilized for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and in whole animal studies [22]. The QSAR approach employs extrathermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, cellular systems, and in vivo. Four standard molecular descriptors routinely used in QSAR analysis are electronic, hydrophobic, steric, and topological indices. The quality of a QSAR model, however, depends mainly on the type and quality of the data, and is valid only for the compound structures analogous to those used to build the model. QSAR models can stand alone, augment other computational approaches, or be examined in tandem with equations of a similar mechanistic genre to establish their authenticity and reliability [23]. Potential use of QSAR models for screening of chemical databases or virtual libraries before their synthesis appears equally attractive to chemical manufacturers, pharmaceutical companies and government agencies.

2. Methods

All the data have been collected from the literature (see individual QSAR for respective references). IC_{50} is the concentration at which the test compounds killed 50% of the mosquito larvae, which is expressed in molar concentration. $\log 1/IC_{50}$ is the subsequent dependent-variable that defines the biological parameter for QSAR development. Physicochemical descriptors are auto-loaded, and multi-regression analysis (MRA) is used to derive the QSAR by utilizing the C-QSAR program [24]. Selection of descriptors is made on the basis of permutation and correlation matrices among the descriptors in order to avoid collinearity problems. Details about the C-QSAR program, the search engine, the choice of parameters and their use in the development of QSAR models, have already been discussed [25,26].

The parameters used in this paper have already been discussed in detail along with their application [19]. Briefly, $\log P$ is the calculated partition coefficient of a compound in *n*-octanol/water and is a measure of its hydrophobicity, while π is the hydrophobic parameter for substituents only. σ, σ^+ and σ^- are Hammett electronic parameters that apply to substituent effects on aromatic systems. E_S is Taft's steric parameter for substituents. MR is the calculated molar refractivity for the whole molecule. Molar refractivity (MR) is calculated from the Lorentz–Lorenz equation and is described as follows: $[(n^2-1)/(n^2+2)](MW/\delta)$, where n is the refractive index, MW is the molecular weight, and δ is the density of the substance. MR is dependent on volume and polarizability. It can be used for a substituent or for the whole molecule. The indicator variable I is assigned the value of 1 or 0 for special features with special effects that cannot be parametrized and has been explained wherever used.

In QSAR equations, n is the number of data points, r is the correlation coefficient between observed values of the dependent and the values calculated from the equation, r^2 is the square of the correlation coefficient and represents the goodness of fit, q^2 is the cross-validated r^2 (a measure of the quality of the QSAR model), and s is the standard deviation. The cross-validated r^2 (q^2) is obtained by using leave-one-out (LOO) procedure as described by Cramer et al. [27]. Q is the quality factor (quality ratio), $Q = r/s$, where r is the correlation coefficient and s is the standard deviation. Chance correlation, due to the excessive number of parameters (which increases the r and s values also), can, thus, be detected by the examination of the Q value. High values of Q indicate the high predictive power of the QSAR models and the lack of "overfitting". F is the Fischer statistics (Fischer ratio), $F = fr^2/[(1-r^2)m]$, where f is the number of degrees of freedom, $f = n-(m+1)$, n is the number of data points, and m is the number of variables. The F-value is actually the ratio between explained and unexplained variance for a given number of degrees of freedom. Thus, it indicates a true relationship, or the significance level for the MLR models. The modeling was taken to be optimal when Q reached a maximum together with F, even if slightly non-optimal F-values have normally been accepted. A significant decrease in F with the introduction of one additional variable (with increasing Q and decreasing s) could mean that the new descriptor is not as significant as expected, that is, its introduction has endangered the statistical quality of the combination. However, the statistical quality could be improved by the introduction of a more convincing descriptor [28–30]. Compounds were deemed to be outliers on the basis of their deviation between observed and calculated activities from the equation ($>2s$) [31–34]. Each regression equation includes 95% confidence limits for each term in parentheses.

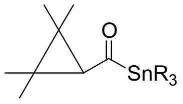
3. Results and discussion

A series of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1** was synthesized and evaluated for their

Table 1
Biological and physicochemical parameters used to derive QSAR Eqs. (1) and (2) for the toxicity of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1** to *Aedes aegypti* and *Anopheles stephensi* mosquito larvae

No.	R	log 1/IC$_{50}$ [Eq. (1)]			log 1/IC$_{50}$ [Eq. (2)]			π_R	MR$_R$
		Obsd.	Pred.	Δ	Obsd.	Pred.	Δ		
1–1	CH_3	5.53	5.49	0.04	4.98	4.90	0.08	0.56	0.56
1–2	C_2H_5	5.45	5.74	−0.29	4.79	5.16	−0.37	1.02	1.03
1–3	*n*-C_3H_7	6.16	6.02	0.14	5.76	5.47	0.29	1.55	1.50
1–4	*n*-C_4H_9	6.33	6.34	−0.01	5.99	5.81	0.18	2.13	1.96
1–5	C_6H_5	6.55	6.25	0.30	5.70	5.71	−0.01	1.96	2.54
1–6	C_6H_{11}	6.53	6.71	−0.18	6.03	6.21	−0.18	2.82	2.67

larvicidal activities against the second instar stage of the *Ae. aegypti*, *An. stephensi*, and *Culex pipiens quinquefasciatus* mosquito larvae by Song et al. [12]. From these data, they developed the following three QSAR models:



1

(i) For the *Ae. aegypti* mosquito larvae, a QSAR was developed between the toxicity of the compounds **1** and the principal moment of inertia along the *z*-axis of the compounds (*Iz*). The equation generated was LC$_{50} = -0.00035 \times (Iz) + 1.74$ with a multiple R^2 of 0.79 and a cross-validation of 0.50. LC$_{50}$ is the concentration at which the test compounds killed 50% of the mosquito larvae, which is expressed in ppm.

(ii) For the *An. stephensi* mosquito larvae, the toxicity of the compounds **1** was correlated to the dipole moment of the molecule along the *x*-axis (P_x). The QSAR equation generated was LC$_{50} = 6.03 \times (P_x) + 0.244$ with a multiple R^2 of 0.92 and a cross-validation of 0.095. In this correlation, the authors [12] were unable to accommodate the ethyl derivative of the complex (**1**, R = C_2H_5) and this compound was considered to be an outlier.

(iii) For the *Cx. pipiens quinquefasciatus* mosquito larvae, the toxicity of the compounds **1** was correlated to the surface area of the molecule. The QSAR equation generated was LC$_{50} = -0.007 \times (\text{surface area}) + 2.52$ with a multiple R^2 of 0.92 and a cross-validation of 0.095.

All the above three QSAR equations were obtained by using different descriptors of the molecules suggesting that the interactions between the molecules and larvae are different for each species. On the basis of these observations the authors [12] suggested that the kill mechanism is different for each species of larvae.

The advantage of using triorganotins as a potential larvicide lies in their ability to biodegrade in the environment to non-toxic tin species. Further, these three species of mosquitos have no reported resistance towards triorganotins [12,35]. These facts encourage us to reevaluate the toxicity data of organotins **1** against *Ae. aegypti* and *An. stephensi* mosquito larvae, to find out any possible similar kill mechanism, and also about the contribution of hydrophobic descriptors of R-groups [Eqs. (1) and (2)].

*3.1. Toxicity of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1** to Ae. aegypti mosquito larvae. Data obtained from Song et al. [12]*

$$\log 1/IC_{50} = 0.54(\pm 0.36)\pi_R + 5.19(\pm 0.66)$$
$$n = 6,\ r^2 = 0.811,\ s = 0.237,\ q^2 = 0.568,$$
$$Q = 3.802,\ F_{1,4} = 17.164 \qquad (1)$$

Hydrophobicity of R-groups (π_R) is found to be the single most important parameter for this data set (Table 1). The linear

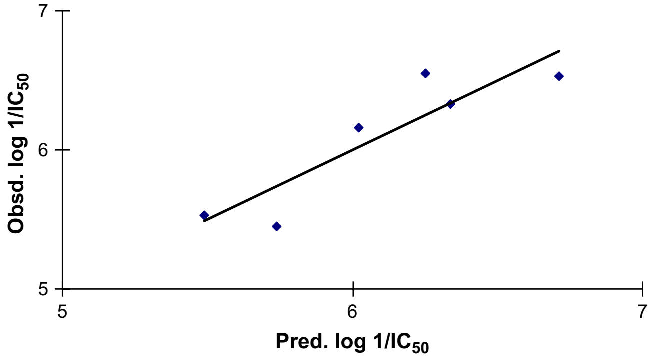


Fig. 1. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (1)].

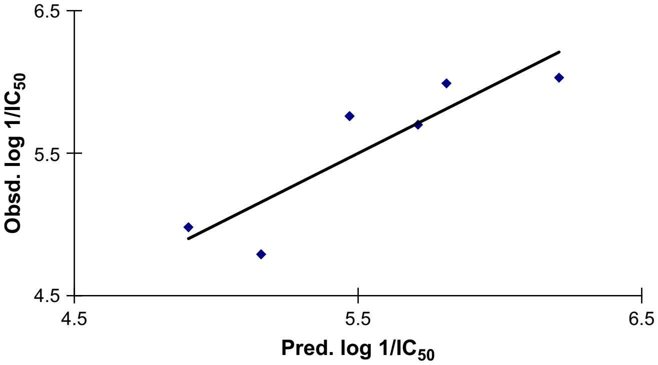


Fig. 2. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (2)].

Table 2
Biological and physicochemical parameters used to derive QSAR Eq. (3) for the toxicity of triphenyltin *para*-substituted benzoates **2** to *Aedes aegypti* mosquito larvae

No.	*X*	log 1/IC$_{50}$ [Eq. (3)]			π_X	MR$_X$
		Obsd.	Pred.	Δ		
2–7	H	5.75	5.76	−0.01	0.00	0.10
2–8	F	5.74	5.73	0.01	0.14	0.09
2–9[a]	Cl	6.23	5.78	0.45	0.71	0.60
2–10	Br	5.87	5.84	0.03	0.86	0.89
2–11	I	5.86	5.95	−0.09	1.12	1.39
2–12	OCH_3	6.10	5.99	0.11	−0.02	0.79
2–13	OH	5.86	5.96	−0.10	−0.67	0.29
2–14	NO_2	6.12	6.02	0.10	−0.28	0.74
2–15	NH_2	6.12	6.16	−0.04	−1.23	0.54
2–16	CH_3	5.78	5.79	−0.01	0.56	0.56
2–17	$C(CH_3)_3$	5.95	5.95	0.00	1.98	1.96

[a] Not used in the derivation of Eq. (3).

π_R model suggests that the compounds with highly hydrophobic R-groups will be more active. A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **1** used in the development of QSAR Eq. (1) is shown in Fig. 1.

It is interesting to note here that there is a high mutual correlation between π_R and MR$_R$ ($r = 0.944$). Where, MR$_R$ is the molar refractivity of R-groups. Thus, it is very hard to predict for this data set if it is a positive hydrophobic or polarizability effect of the R-groups. We derived Eq. (1a) in term of MR$_R$.

$$\log 1/\mathrm{IC}_{50} = 0.55(\pm 0.27)\mathrm{MR}_\mathrm{R} + 5.15(\pm 0.50)$$
$$n = 6,\ r^2 = 0.891,\ s = 0.180,\ q^2 = 0.781,$$
$$Q = 5.244,\ F_{1,4} = 32.697 \qquad (1a)$$

Although, statistics of Eq. (1a) are better than those of Eq. (1), we preferred Eq. (1) because the descriptor π_R is more informative than that of MR$_R$. Molar refractivity (MR$_R$) is the parameter to measure the combined effect of bulk and polarizability of R-groups.

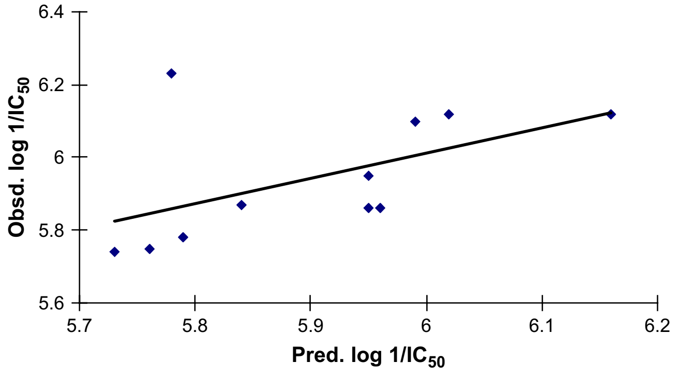


Fig. 3. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (3)].

*3.2. Toxicity of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1** to An. stephensi mosquito larvae. Data obtained from Song et al. [12]*

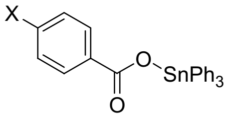
$$\log 1/\mathrm{IC}_{50} = 0.58(\pm 0.41)\pi_\mathrm{R} + 4.57(\pm 0.76)$$
$$n = 6,\ r^2 = 0.792,\ s = 0.270,\ q^2 = 0.538,$$
$$Q = 3.296,\ F_{1,4} = 15.231 \quad \pi_\mathrm{R}\ \text{versus}\ \mathrm{MR}_\mathrm{R},$$
$$r = 0.944 \qquad (2)$$

$$\log 1/\mathrm{IC}_{50} = 0.52(\pm 0.50)\mathrm{MR}_\mathrm{R} + 4.65(\pm 0.93)$$
$$n = 6,\ r^2 = 0.682,\ s = 0.333,\ q^2 = 0.408,$$
$$Q = 2.480,\ F_{1,4} = 8.579 \qquad (2a)$$

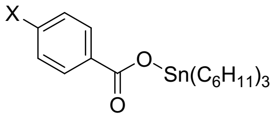
We preferred Eq. (2) on the basis of the above reason as well as on their statistics, which are better than those of Eq. (2a) (Table 1). It is also interesting to note here that Eq. (2a) is statistically not significant. A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **1** used in the development of QSAR Eq. (2) is shown in Fig. 2.

Reasonable QSARs (1) and (2) with correlation coefficients $r^2 = 0.811$ and 0.792 were obtained for the toxicities of organotins **1** to *Ae. aegypti* and *An. stephensi* mosquito larvae with π_R (hydrophobicity of R-groups). However, owing to the limited number of compounds used in generating the QSARs, care should be taken in their use. QSAR Eq. (1) is very similar to that of (2) expressing the similar kill mechanism for both species of larvae, which is contradictory to the findings of Song et al. [12].

Duong et al. [14] synthesized two series of organotin compounds that are triphenyltin *para*-substituted benzoates **2** and tricyclohexyltin *para*-substituted benzoates **3**, and evaluated their larvicidal activities against the second instar stage of the *Ae. aegypti* and *An. stephensi* mosquito larvae.



2



3

The authors [14] failed to generate a QSAR model from the toxicity data of these compounds against *Ae. aegypti* mosquito larvae but succeeded against that of *An. stephensi*. The QSAR was generated using the formula weights (fw) and the valence 3rd path χ index (xvp3) of the molecules. The equation was $\mathrm{LC}_{50} = 0.0120(\mathrm{fw}) - 0.308(\mathrm{xvp3}) + 2.52$ with a multiple R^2 of 0.78 and a cross-validation of 15.8. On the basis of this finding, the authors [14] suggested that the kill mechanism is different for these two species of mosquitoes.

We are interested to evaluate the contribution of *X*-substituents on the kill activity of organotins **2** and **3** against *Ae. aegypti* and *An. stephensi* mosquito larvae, and developed three QSAR models [(3)–(5)]. One QSAR for the toxicity of compounds **2** against *An. stephensi* mosquito larvae was not

Table 3
Biological and physicochemical parameters used to derive QSAR Eqs. (4) and (5) for the toxicity of tricyclohexyltin *para*-substituted benzoates **3** to *Aedes aegypti* and *Anopheles stephensi* mosquito larvae

No.	X	log 1/IC$_{50}$ [Eq. (4)]			log 1/IC$_{50}$ [Eq. (5)]			π_X	E_{S-X}	σ_X^+
		Obsd.	Pred.	Δ	Obsd.	Pred.	Δ			
3–18	H	5.73	5.74	−0.01	5.87	5.76	0.11	0.00	0.00	0.00
3–19	F	5.55	5.70	−0.15	5.55	5.82	−0.27	0.14	−0.55	−0.07
3–20	Cl	5.74	5.77	−0.03	5.97	5.67	0.30	0.71	−0.97	0.11
3–21	Br	5.80	5.78	0.02	5.29	5.64	−0.35	0.86	−1.16	0.15
3–22	I	5.86	5.77	0.09	5.68	5.65	0.03	1.12	−1.62	0.14
3–23[a]	OCH_3	5.99	5.66	0.33	6.42	6.40	0.02	−0.02	−0.55	−0.78
3–24	OH	5.54	5.52	0.02	6.43	6.52	−0.09	−0.67	−0.55	−0.92
3–25[b]	NO_2	5.29	5.36	−0.07	5.65	5.11	0.54	−0.28	−2.52	0.79
3–26[b]	NH_2	5.49	5.39	0.10	5.59	6.83	−1.24	−1.23	−0.61	−1.30
3–27	CH_3	5.70	5.70	0.00	6.14	6.02	0.12	0.56	−1.24	−0.31
3–28	$C(CH_3)_3$	5.82	5.81	0.01	6.10	5.98	0.12	1.98	−2.78	−0.26

[a] Not used in the derivation of Eq. (4).
[b] Not used in the derivation of Eq. (5).

formulated. This may be not possible due to a very low variation in the values of their log 1/IC$_{50}$.

3.3. *Toxicity of triphenyltin para-substituted benzoates* **2** *to Ae. aegypti mosquito larvae. Data obtained from Duong et al. [14]*

$$\log 1/\mathrm{IC}_{50} = -0.21(\pm 0.10)\pi_X + 0.32(\pm 0.16)\mathrm{MR}_X + 5.73(\pm 0.12)$$
$$n = 10,\ r^2 = 0.789,\ s = 0.078,\ q^2 = 0.646,\ Q = 11.385,\ F_{2,7} = 13.088 \quad \text{outlier}: X = \mathrm{Cl} \qquad (3)$$

π_X and MR$_X$ are the hydrophobic and molar refractivity parameters of X-substituents (Table 2). Negative coefficient of π_X suggests that the toxicity of these compounds decreases with the increase of hydrophobicity of X-substituents. On the other hand, an increase in the molar refractivity of X-substituents increases the toxicity of these compounds (positive coefficient). One compound (**2–9**, $X = \mathrm{Cl}$) was not used in the derivation of Eq. (3) due to the high deviation between the observed and predicted toxicity(Obsd. − Pred. = 6.23 − 5.78 = 0.45 > 5 × s). A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **2** used in the development of QSAR Eq. (3) is shown in Fig. 3.

3.4. *Toxicity of tricyclohexyltin para-substituted benzoates* **3** *to Ae. aegypti mosquito larvae. Data obtained from Duong et al. [14]*

$$\log 1/\mathrm{IC}_{50} = 0.22(\pm 0.08)\pi_X + 0.13(\pm 0.08)E_{S-X} + 5.74(\pm 0.11)$$
$$n = 10,\ r^2 = 0.849,\ s = 0.080,\ q^2 = 0.592,\ Q = 11.513,\ F_{2,7} = 19.679 \quad \text{outlier}: X = \mathrm{OCH_3} \qquad (4)$$

Positive π_X suggests that the toxicity of these compounds increases with the increase of hydrophobicity of X-substituents (Table 3). E_{S-X} is Taft's steric parameter for X-substituents. Since the values of E_{S-X} constants are all negative, the positive coefficient with this term in Eq. (4) indicates a small steric effect. In other words, the smaller the substituent, the stronger the toxicity is. This might be due to polarizability of the

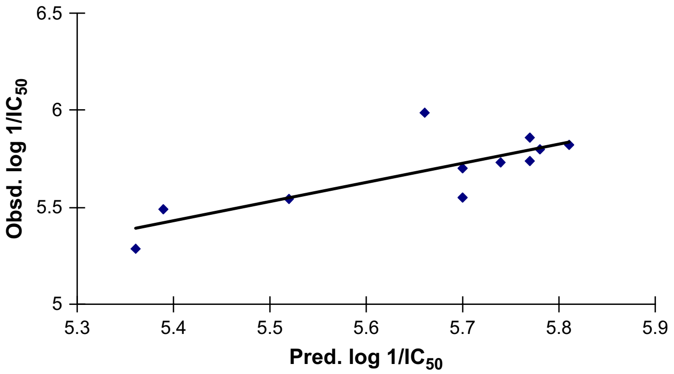


Fig. 4. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (4)].

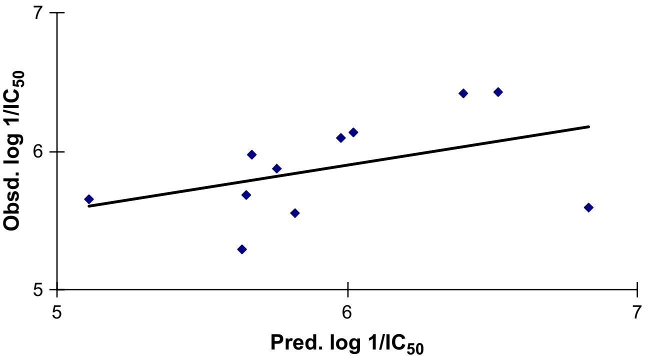
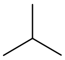
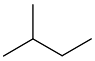
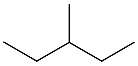
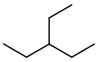


Fig. 5. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (5)].

Table 4
knotp values of the compound's fragment containing xc3 and xpc4 subgraphs

No.	Fragment	Description		xc3	xpc4	knotp
1		1 Cluster-3	0 Path cluster-4	0.5774	0.0000	0.5774
2		1 Cluster-3	1 Path cluster-4	0.4082	0.4082	0.0000
3		1 Cluster-3	2 Path cluster-4	0.2887	0.5774	−0.2887
4		1 Cluster-3	3 Path cluster-4	0.2041	0.6124	−0.4083

X-substituent. One compound (**3–23**, $X = OCH_3$) was not used in the derivation of Eq. (4) due to the high deviation between the observed and predicted toxicity (Obsd. − Pred. $= 5.99 - 5.66 = 0.33 > 4 \times s$). A comparison between observed and predicted log $1/IC_{50}$ of organotin derivatives **3** used in the development of QSAR Eq. (4) is shown in Fig. 4.

3.5. Toxicity of tricyclohexyltin para-substituted benzoates 3 to An. stephensi mosquito larvae. Data obtained from Duong et al. [14]

$$\begin{aligned}\log 1/IC_{50} = {} & -0.82(\pm 0.46)\sigma_X^+ + 5.76(\pm 0.20) \\ & n = 9,\ r^2 = 0.719,\ s = 0.218,\ q^2 = 0.565, \\ & Q = 3.890,\ F_{1,7} = 17.911 \quad \text{outliers :} \\ & X = NO_2;\ NH_2 \end{aligned} \tag{5}$$

σ_X^+ is the values of σ^+ of *X*-substituents (Table 3). The negative coefficient with σ_X^+ (−0.82) implies that highly electron releasing substituents at position *X* (e.g. $NHCH_3$, $N(CH_3)_3$, etc.) will enhance the kill activity of compounds **3** to *An. stephensi* mosquito larvae. Two compounds (**3–25**, $X = NO_2$ and **3–26**, $X = NH_2$) were not used in the development of Eq. (5). The outlier (**3–25**, $X = NO_2$) is much more active than the expected by 2.5 times the standard deviation. This may be due to the formation of the nitro anion radical that interact with DNA [20]. The other derivative (**3–26**, $X = NH_2$) was also considered to be an outlier due to being much less active than expected by 5.7 times the standard deviation. This anomalous behavior may be attributed due to its nature as aniline. This could result in the hydrogen abstraction [20,36]. A comparison between observed and predicted log $1/IC_{50}$ of organotin derivatives **3** used in the development of QSAR Eq. (5) is shown in Fig. 5.

Eqs. (4) and (5) for the kill activities of tricyclohexyltin *para*-substituted benzoates **3** are very different to each other suggest that these compounds may act by a different kill mechanism for these two species of larvae (*Ae. aegypti* and *An. stephensi* mosquito larvae), which is also supported by the findings of Duong et al. [14].

A series of tris-(*para*-substituted phenyl)tins **4** was also synthesized by Song et al. [16] and evaluated for their larvicidal activities against the second instar stage of the *Ae.*

Table 5
Biological and physicochemical parameters used to derive QSAR Eqs. (6) and (7) for the toxicity of tris-(*para*-substituted phenyl)tins **4** to *Aedes aegypti* and *Anopheles stephensi* mosquito larvae

No.	*X*	*Y*	log $1/IC_{50}$ [Eq. (6)]			log $1/IC_{50}$ [Eq. (7)]			π_X	σ_X^+
			Obsd.	Pred.	Δ	Obsd.	Pred.	Δ		
4–29	H	Cl	5.18	5.27	−0.09	6.19	6.13	0.06	0.00	0.00
4–30	H	OH	5.39	5.27	0.12	6.42	6.13	0.29	0.00	0.00
4–31[a]	H	OAc	5.25	5.27	−0.02	7.01	6.13	0.88	0.00	0.00
4–32	CH_3	Cl	5.61	5.64	−0.03	5.09	5.24	−0.15	0.56	−0.31
4–33	CH_3	OH	5.72	5.64	0.08	5.22	5.24	−0.02	0.56	−0.31
4–34	CH_3	OAc	5.58	5.64	−0.06	5.01	5.24	−0.23	0.56	−0.31
4–35	F	Cl	5.72	5.90	−0.18	5.80	5.92	−0.12	0.14	−0.07
4–36	F	OH	5.93	5.90	0.03	5.71	5.92	−0.21	0.14	−0.07
4–37	F	OAc	6.05	5.90	0.15	5.87	5.92	−0.05	0.14	−0.07
4–38	Cl	Cl	5.48	5.54	−0.06	5.75	5.69	0.06	0.71	0.11
4–39	Cl	OH	5.64	5.54	0.10	5.66	5.69	−0.03	0.71	0.11
4–40	Cl	OAc	5.48	5.54	−0.06	5.81	5.69	0.12	0.71	0.11
4–41	SCH_3	Cl	ND	5.01	ND	4.79	4.81	−0.02	0.61	−0.60
4–42	SCH_3	OH	ND	5.01	ND	5.00	4.81	0.19	0.61	−0.60
4–43	SCH_3	OAc	ND	5.01	ND	4.91	4.81	0.10	0.61	−0.60

ND = not determined.

[a] Not used in the derivation of Eq. (7).

aegypti and *An. stephensi* mosquito larvae. From these data, they developed the following QSAR models:

$$(4\text{ - X} - \text{C}_6\text{H}_4)_3\text{SnY}$$

4

(i) For the *Ae. aegypti* mosquito larvae, a QSAR was developed between the toxicity of the compounds **4** and two descriptors of the molecules, the kappa shape index ($k1$) and kappa alpha shape index (ka2), which is a modified version of the kappa shape index. Both of these indices are attributes related to the molecular shape encoded in the molecules. The QSAR generated was $\text{LC}_{50} = 2.388 \times \text{ka2} - 0.9808 \times k1 + 2.56986$ with a multiple R^2 of 0.7379 and a cross-validation of 4.36. During the formulation of this QSAR, three compounds (**4–41**, **4–42**, and **4–43**) were excluded due to uncertain results.

(ii) In the case of *An. stephensi* mosquito larvae, a QSAR was also developed between the toxicity of the compounds **4** and two descriptors of the molecules, the ovality and knotp. The ovality is equal to surface/$4\pi r^2$, where $r = (3 \times \text{volume}/4\pi)^{1/3}$ and surface is defined as the surface area. The knotp descriptor gives the difference between chi cluster-3 and chi path/cluster-4 descriptor. Thus, knotp is defined as: knotp = xc3 − xpc4. The value of knotp is largest where an xc3 subgraph is not associated with an xpc4 subgraph. The knotp descriptor will differentiate between xc3 and xpc4 in situations where xc3 and xpc4 have the same value. Increased adjacent branching of the xc3 subgraph causes an increase xpc4 value resulting in a decrease in the value of knotp. When the adjacent branching involves hetero atoms the decrease in knotp is lessened because the value of xpc4 subgraphs is reduced by the presence of heteroatom. Every path cluster-4 (xpc4) subgraph contains a cluster-3 (xc3) subgraph and one additional atom. Each xc3 subgraph may be associated with up to three of these additional atoms and thus be contained within up to three xpc4 subgraphs, as shown in Table 4. The knotp descriptor helps to separate this overlapping structure information into distinct numerical values [37]. The QSAR generated was $\text{LC}_{50} = -5.351 \times \text{knotp} + 25.52 \times \text{ovality} - 50.555$ with a multiple R^2 of 0.830 and a cross-validation of 28.61. In this model, all the 15 compounds were used.

On the basis of the above QSAR equations, authors [16] suggested that the size of the substituent on the ring is responsible for their toxicity. Now, we are interested to find out any possibility for the contribution of hydrophobicity and electronic effects of the substituents *X* and *Y*. Thus, we developed two QSAR Eqs. (6) and (7).

3.6. Toxicity of tris-(*para*-substituted phenyl)tins **4** to *Ae. aegypti* mosquito larvae. Data obtained from Song et al. [16]

$$\begin{aligned}\log 1/\text{IC}_{50} &= 6.46(\pm 2.32)\pi_X - 8.91(\pm 3.23)\pi_X^2 \\ &\quad + 1.49(\pm 0.86)\sigma_X^+ + 5.28(\pm 0.15) \\ &n = 12,\ r^2 = 0.851,\ s = 0.115,\ q^2 = 0.664, \\ &Q = 8.017,\ F_{3,8} = 15.230 \quad \text{optimum} \\ &\pi_X = 0.36(0.35 - 0.38) \quad \pi_X \text{ versus } \sigma_X^+,\ r = 0.056 \qquad (6)\end{aligned}$$

It is a parabolic correlation in terms of π_X (hydrophobicity of *X*-substituents) followed by linearly related to σ_X^+ ($\sigma^+$of *X*-substituents), which suggests that the kill activities of compounds **4** against *Ae. aegypti* mosquito larvae first increases with an increase in hydrophobicity of *X*-substituents up to an optimum value of $\pi_X = 0.36$ and then decreases (Table 5). The positive coefficient with σ_X^+ (1.49) implies that highly electron withdrawing substituents at position *X* (e.g. NO_2, CN, CHO, CF_3, $COCF_3$, etc.) will enhance the kill activity of the compounds. No contribution of *Y*-substituents was found. Thus, one should increase the toxicity of compounds **4** to *Ae. aegypti* mosquito larvae by boosting the electron withdrawing nature of *X*-substituents and preserving $\pi_X \approx 0.36$. A comparison between observed and predicted log $1/\text{IC}_{50}$ of organotin derivatives **4** used in the development of QSAR Eq. (6) is shown in Fig. 6.

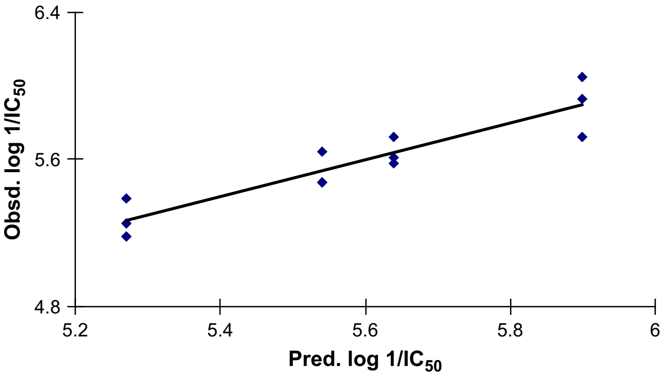


Fig. 6. Plot of observed versus predicted log $1/\text{IC}_{50}$ [Eq. (6)].

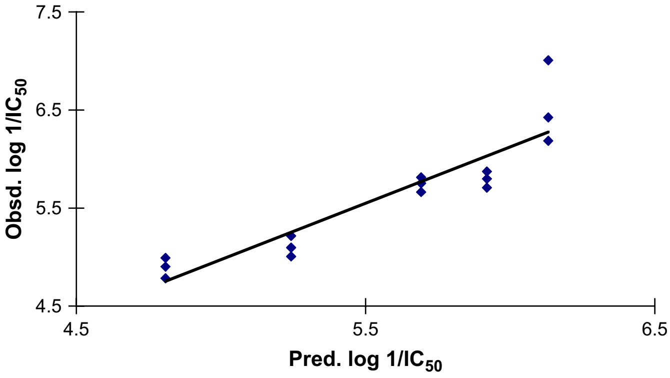


Fig. 7. Plot of observed versus predicted log $1/\text{IC}_{50}$ [Eq. (7)].

Table 6

Biological, physicochemical, and structural parameters used to derive QSAR Eqs. (8) and (9) for the toxicity of triorganotin 2,2,3,3-tetramethylcyclopropane-carboxylates **1**, triphenyltin *para*-substituted benzoates **2**, tricyclohexyltin *para*-substituted benzoates **3**, and tris-(*para*-substituted phenyl)tins **4** to *Aedes aegypti* and *Anopheles stephensi* mosquito larvae

No.	$\log 1/IC_{50}$ [Eq. (8)]			$\log 1/IC_{50}$ [Eq. (9)]			$\log P$	MR	I_1	I_3	I_4
	Obsd.	Pred.	Δ	Obsd.	Pred.	Δ					
1–1[b]	5.53	5.53	0.00	4.98	7.11	−2.13	1.68	6.94	1	0	0
1–2[b]	5.45	5.76	−0.31	4.79	6.81	−2.02	3.27	8.33	1	0	0
1–3[b]	6.16	5.99	0.17	5.76	6.51	−0.75	4.86	9.73	1	0	0
1–4	6.33	6.23	0.10	5.99	6.21	−0.22	6.45	11.12	1	0	0
1–5	6.55	6.25	0.30	5.70	5.79	−0.09	6.64	13.09	1	0	0
1–6	6.53	6.42	0.11	6.03	5.73	0.30	7.78	13.37	1	0	0
2–7[c]	5.75	5.94	−0.19	5.19	5.28	−0.09	4.47	12.49	0	0	0
2–8[c]	5.74	5.96	−0.22	5.09	5.27	−0.18	4.62	12.50	0	0	0
2–9[c]	6.23	6.04	0.19	5.21	5.17	0.04	5.19	12.98	0	0	0
2–10[c]	5.87	6.06	−0.19	5.25	5.11	0.14	5.34	13.26	0	0	0
2–11[c]	5.86	6.10	−0.24	5.10	5.00	0.10	5.60	13.79	0	0	0
2–12[c]	6.10	5.93	0.17	5.13	5.15	−0.02	4.39	13.10	0	0	0
2–13[c]	5.86	5.84	0.02	5.40	5.24	0.16	3.81	12.64	0	0	0
2–14[c]	6.12	5.90	0.22	5.05	5.15	−0.10	4.22	13.10	0	0	0
2–15[c]	6.12	5.76	0.36	5.26	5.20	0.06	3.25	12.86	0	0	0
2–16[c]	5.78	6.01	−0.23	5.00	5.18	−0.18	4.97	12.95	0	0	0
2–17[c]	5.95	6.20	−0.25	5.12	4.88	0.24	6.30	14.34	0	0	0
3–18	5.73	5.61	0.12	5.87	6.18	−0.31	4.47	12.49	0	1	0
3–19[b]	5.55	5.63	−0.08	5.55	6.18	−0.63	4.62	12.50	0	1	0
3–20	5.74	5.71	0.03	5.97	6.07	−0.10	5.19	12.98	0	1	0
3–21[b]	5.80	5.73	0.07	5.29	6.01	−0.72	5.34	13.26	0	1	0
3–22	5.86	5.77	0.09	5.68	5.90	−0.22	5.60	13.79	0	1	0
3–23[a]	5.99	5.60	0.39	6.42	6.05	0.37	4.39	13.10	0	1	0
3–24	5.54	5.51	0.03	6.43	6.15	0.28	3.81	12.64	0	1	0
3–25	5.29	5.57	−0.28	5.65	6.05	−0.40	4.22	13.10	0	1	0
3–26[b]	5.49	5.43	0.06	5.59	6.10	−0.51	3.25	12.86	0	1	0
3–27	5.70	5.68	0.02	6.14	6.08	0.06	4.97	12.95	0	1	0
3–28	5.82	5.87	−0.05	6.10	5.78	0.32	6.30	14.34	0	1	0
4–29	5.18	5.35	−0.17	6.19	5.85	0.34	3.57	9.82	0	0	1
4–30	5.39	5.34	0.05	6.42	5.92	0.50	3.51	9.48	0	0	1
4–31[b]	5.25	5.33	−0.08	7.01	5.72	1.29	3.46	10.44	0	0	1
4–32	5.61	5.57	0.04	5.09	5.55	−0.46	5.06	11.21	0	0	1
4–33	5.72	5.56	0.16	5.22	5.62	−0.40	5.00	10.87	0	0	1
4–34	5.58	5.55	0.03	5.01	5.42	−0.41	4.96	11.83	0	0	1
4–35	5.72	5.41	0.31	5.80	5.84	−0.04	3.99	9.86	0	0	1
4–36[a]	5.93	5.40	0.53	5.71	5.91	−0.20	3.93	9.52	0	0	1
4–37[a]	6.05	5.39	0.66	5.87	5.71	0.16	3.89	10.49	0	0	1
4–38	5.48	5.66	−0.18	5.75	5.53	0.22	5.70	11.29	0	0	1
4–39	5.64	5.65	−0.01	5.66	5.61	0.05	5.64	10.95	0	0	1
4–40	5.48	5.64	−0.16	5.81	5.40	0.41	5.60	11.91	0	0	1
4–41	ND	5.59	ND	4.79	5.03	−0.24	5.24	13.63	0	0	1
4–42	ND	5.58	ND	5.00	5.10	−0.10	5.18	13.29	0	0	1
4–43	ND	5.58	ND	4.91	4.90	0.01	5.14	14.25	0	0	1

ND = not determined.

[a] Not used in the derivation of Eq. (8).

[b] Not used in the derivation of Eq. (9).

[c] Not used in the development of a separate QSAR for the toxicity of organotins **2** to *An. stephensi* mosquito larvae.

3.7. Toxicity of tris-(para-substituted phenyl)tins 4 to An. stephensi mosquito larvae. Data obtained from Song et al. [16]

$$\log 1/IC_{50} = -0.83(\pm0.37)\pi_X + 1.36(\pm0.39)\sigma_X^+ + 6.13(\pm0.18) \quad n = 14,\ r^2 = 0.914,$$
$$s = 0.162,\ q^2 = 0.862,\ Q = 5.901,\ F_{2,11} = 58.453$$
$$\pi_X \text{ versus } \sigma_X^+,\ r = 0.337 \quad \text{outlier}: X = \mathrm{H},\ \mathrm{Y} = \mathrm{OAc} \tag{7}$$

The negative coefficient with π_X suggests that the highly hydrophilic *X*-substituents will increase the toxicity (Table 5). On the other hand, the positive coefficient with σ_X^+ (1.36) implies that highly electron withdrawing *X*-substituents (e.g. NO_2, CN, CHO, CF_3, $COCF_3$, etc.) will enhance the toxicity of the compounds. Again, no contribution of *Y*-substituents was found. One compound (**4–31**, *X* = H, *Y* = OAc) was not used in the derivation of Eq. (7) due to the high deviation between the observed and predicted toxicity

(Obsd. − Pred. = 7.01 − 6.13 = 0.88 > 5 × s). A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **4** used in the development of QSAR Eq. (7) is shown in Fig. 7.

QSAR Eqs. (6) and (7) for the kill activities of tris-(*para*-substituted phenyl)tins **4** to *Ae. aegypti* and *An. stephensi* mosquito larvae were developed by using two same descriptors (π_X and σ_X^+) but their nature is different from each other, which suggests that these compounds may act by different kill mechanism for these two species of larvae.

Finally, we derived Eqs. (8) and (9) from the combined data of organotin compounds **1–4** for better understanding about their kill mechanism towards the two species of larvae, that are *Ae. aegypti* and *An. stephensi* mosquito larvae.

3.8. Toxicity of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1**, triphenyltin para-substituted benzoates **2**, tricyclohexyltin para-substituted benzoates **3**, and tris-(para-substituted phenyl)tins **4** to Ae. aegypti mosquito larvae [12,14,16]

$$\begin{aligned}\log 1/\mathrm{IC}_{50} = {} & 0.15(\pm 0.05)\log P - 0.33(\pm 0.15)I_3 \\ & - 0.46(\pm 0.15)I_4 + 5.28(\pm 0.27) \\ & n = 37,\ r^2 = 0.711,\ s = 0.184,\ q^2 = 0.645, \\ & Q = 4.582,\ F_{3,33} = 27.062 \\ \text{outliers}: & \ \mathbf{3{-}23},\ \mathbf{4{-}36},\ \mathbf{4{-}37} \end{aligned} \qquad (8)$$

log P is the calculated hydrophobicity for the whole molecules, whereas I_3 is an indicator variable taking the value of 1 or 0 for the presence or absence of compound series **3** (Table 6). Similarly, I_4 is an indicator variable taking the value of 1 or 0 for the presence or absence of compound series **4**. A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **1–4** used in the development of QSAR Eq. (8) is shown in Fig. 8.

3.9. Toxicity of triorganotin 2,2,3,3-tetramethylcyclopropanecarboxylates **1**, triphenyltin para-substituted benzoates **2**, tricyclohexyltin para-substituted benzoates **3**, and tris-(para-substituted phenyl)tins **4** to An. stephensi mosquito larvae [12,14,16]

$$\begin{aligned}\log 1/\mathrm{IC}_{50} = {} & -0.22(\pm 0.07)\mathrm{MR} + 0.64(\pm 0.33)I_1 \\ & + 0.90(\pm 0.23)I_3 + 7.96(\pm 0.85) \\ & n = 36,\ r^2 = 0.722,\ s = 0.265, \\ & q^2 = 0.635,\ Q = 3.208,\ F_{3,32} = 27.703 \\ \text{outliers}: & \ \mathbf{1{-}1},\ \mathbf{1{-}2},\ \mathbf{1{-}3},\ \mathbf{3{-}19},\ \mathbf{3{-}21},\ \mathbf{3{-}26},\ \mathbf{4{-}31} \\ \log P \text{ versus MR}, & \ r = 0.424 \end{aligned} \qquad (9)$$

MR is the calculated molar refractivity for the whole molecules, whereas I_1 is an indicator variable taking the value of 1 or 0 for the presence or absence of compound series **1** (Table 6). Similarly, I_3 is an indicator variable taking the value of 1 or 0 for the presence or absence of compound series **3**.

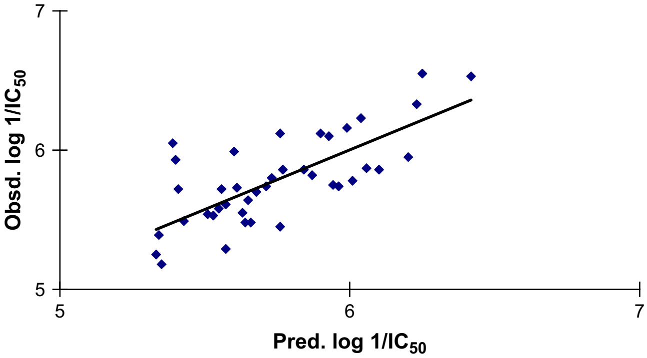


Fig. 8. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (8)].

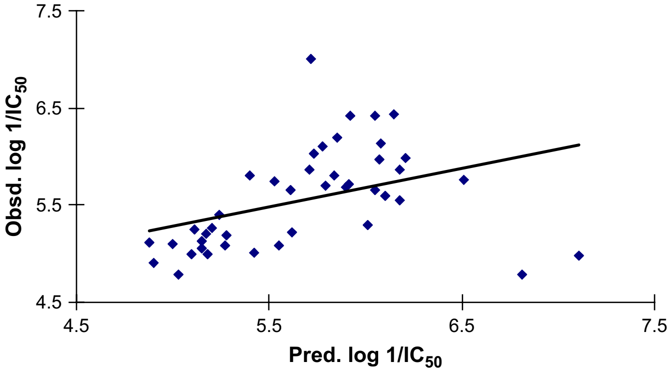


Fig. 9. Plot of observed versus predicted log 1/IC$_{50}$ [Eq. (9)].

A comparison between observed and predicted log 1/IC$_{50}$ of organotin derivatives **1–4** used in the development of QSAR Eq. (9) is shown in Fig. 9.

The combined Eqs. (8) and (9) are actually dominated by the indicator variables. This may be due to the fact that these equations were developed by the toxicity data of four different series of compounds, where we can use only the descriptors for the whole molecules and not for their substituents. Eqs. (8) and (9) for the kill activities of organotins **1–4** to *Ae. aegypti* and *An. stephensi* mosquito larvae, respectively, are very different to each other suggests that these compounds may act by very different kill mechanism for these two species of larvae.

4. Validation of the QSAR models

QSAR model validation becomes an essential part in developing a statistically valid and predictive model, because the real utility of a QSAR model is in its ability to predict accurately the modeled property for new compounds. The following approaches have been used to validate the QSAR Eqs. [(1)–(9)].

4.1. Internal validation

- *Fraction of the variance*: The fraction of the variance of an MRA model is expressed by r^2. It is believed that the closer

Table 7
Y-randomization test data for Eqs. [(1)–(9)]

Eq. no.	NOR-1		NOR-2		NOR-3		NOR-4		NOR-5	
	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2	r^2	q^2
(1)	0.081	−1.830	0.132	−1.279	0.248	−0.753	0.248	−0.663	0.028	−2.746
(2)	0.022	−1.410	0.008	−1.375	0.171	−1.082	0.397	−0.334	0.009	−2.712
(3)	0.379	−0.035	0.089	−1.179	0.087	−0.761	0.091	−1.064	0.338	−0.087
(4)	0.096	−1.581	0.049	−1.555	0.000	−1.582	0.123	−1.322	0.319	−0.816
(5)	0.041	−1.207	0.426	−0.103	0.006	−0.555	0.029	−0.510	0.024	−0.344
(6)	0.282	−0.449	0.223	−0.424	0.354	−1.696	0.448	−0.094	0.301	−1.050
(7)	0.270	−0.068	0.196	−0.453	0.140	−0.585	0.221	−0.087	0.008	−0.516
(8)	0.440	0.305	0.441	0.313	0.443	0.313	0.445	0.314	0.440	0.310
(9)	0.405	0.277	0.435	0.335	0.412	0.274	0.432	0.321	0.391	0.285

NOR = number of Y-randomization.

the value of r^2 to unity, the better the QSAR model. The values of r^2 for these QSAR models are from 0.711 to 0.914, which suggests that these QSAR models explain 71.1–91.4% of the variance of the data. According to the literature, the predictive QSAR model must have $r^2 > 0.6$ [38,39].

- *Cross-validation test*: The values of q^2 for these QSAR models are from 0.538 to 0.862. The high values of q^2 validate the QSAR models. According to the literature, the predictive QSAR model must have $q^2 > 0.5$ [38,39].
- *Standard deviation* (s): s is the standard deviation about the regression line. The smaller the value of s the better the QSAR model. The values of s for these QSAR models are 0.078–0.270.
- *Quality factor* (Q): Chance correlation, due to the excessive number of parameter (which increases also the r and s values) is detected by the examination of Q value [28–30]. The high values of Q (3.208–11.513) for these QSAR models suggest their high predictive power and lack of overfitting.
- *Fischer statistics* (F): Fischer statistics (F) is the ratio between explained and unexplained variance for a given number of degrees of freedom. The larger the value of F the greater the probability that the QSAR equation is significant. The F-values for these QSAR models are obtained from 13.088 to 58.453, which are statistically significant at the 95% level.
- *Y-randomization test*: In this test, the dependent-variable vector (Y-vector) is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The process is repeated several times. It is expected that the resulting QSAR models should have low r^2 and low

Table 8
QSAR obtained by the use of fewer numbers of descriptors than that of the original QSAR with their statistical parameters

OEN	NSQ	QSAR with fewer numbers of descriptors than that of the original QSAR	n	r^2	q^2	s
(3)	3a	$\log 1/\mathrm{IC}_{50} = -0.05(\pm 0.13)\pi_X + 5.93(\pm 0.11)$	10	0.107	−0.470	0.150
	3b	$\log 1/\mathrm{IC}_{50} = 0.08(\pm 0.20)\mathrm{MR}_X + 5.86(\pm 0.19)$	10	0.087	−0.312	0.152
(4)	4a	$\log 1/\mathrm{IC}_{50} = 0.15(\pm 0.10)\pi_X + 5.61(\pm 0.10)$	10	0.584	0.374	0.124
	4b	$\log 1/\mathrm{IC}_{50} = 0.001(\pm 0.17)E_{S-X} + 5.66(\pm 0.24)$	10	0.000	−0.992	0.192
(6)	6a	$\log 1/\mathrm{IC}_{50} = 0.10(\pm 0.58)\pi_X + 5.55(\pm 0.27)$	12	0.016	−0.404	0.264
	6b	$\log 1/\mathrm{IC}_{50} = -0.46(\pm 1.07)\sigma_X^+ + 5.56(\pm 0.18)$	12	0.083	−0.141	0.255
	6c	$\log 1/\mathrm{IC}_{50} = 0.09(\pm 0.60)\pi_X - 0.45(\pm 1.13)\sigma_X^+ + 5.53(\pm 0.28)$	12	0.095	−0.340	0.267
	6d	$\log 1/\mathrm{IC}_{50} = 3.19(\pm 2.16)\pi_X - 4.35(\pm 2.99)\pi_X^2 + 5.37(\pm 0.23)$	12	0.553	0.235	0.188
(7)	7a	$\log 1/\mathrm{IC}_{50} = -1.22(\pm 0.86)\pi_X + 6.04(\pm 0.44)$	14	0.442	0.242	0.395
	7b[a]	$\log 1/\mathrm{IC}_{50} = 1.62(\pm 0.63)\sigma_X^+ + 5.82(\pm 0.20)$	14	0.725	0.642	0.277
(8)	8a	$\log 1/\mathrm{IC}_{50} = 0.16(\pm 0.08)\log P + 5.01(\pm 0.39)$	37	0.325	0.229	0.274
	8b	$\log 1/\mathrm{IC}_{50} = -0.16(\pm 0.24)I_3 + 5.81(\pm 0.13)$	37	0.049	−0.037	0.325
	8c	$\log 1/\mathrm{IC}_{50} = -0.36(\pm 0.22)I_4 + 5.87(\pm 0.11)$	37	0.248	0.177	0.289
	8d	$\log 1/\mathrm{IC}_{50} = 0.16(\pm 0.08)\log P - 0.16(\pm 0.20)I_3 + 5.05(\pm 0.38)$	37	0.373	0.268	0.268
	8e	$\log 1/\mathrm{IC}_{50} = 0.15(\pm 0.07)\log P - 0.34(\pm 0.17)I_4 + 5.14(\pm 0.33)$	37	0.535	0.454	0.230
	8f	$\log 1/\mathrm{IC}_{50} = -0.34(\pm 0.20)I_3 - 0.49(\pm 0.20)I_4 + 6.00(\pm 0.12)$	37	0.440	0.352	0.253
(9)	9a	$\log 1/\mathrm{IC}_{50} = -0.12(\pm 0.11)\mathrm{MR} + 7.07(\pm 1.42)$	36	0.121	0.023	0.457
	9b	$\log 1/\mathrm{IC}_{50} = 0.38(\pm 0.58)I_1 + 5.52(\pm 0.17)$	36	0.050	−0.020	0.475
	9c	$\log 1/\mathrm{IC}_{50} = 0.61(\pm 0.33)I_3 + 5.42(\pm 0.16)$	36	0.289	0.218	0.411
	9d	$\log 1/\mathrm{IC}_{50} = -0.13(\pm 0.11)\mathrm{MR} + 0.40(\pm 0.55)I_1 + 7.07(\pm 1.39)$	36	0.177	0.065	0.449
	9e	$\log 1/\mathrm{IC}_{50} = -0.20(\pm 0.08)\mathrm{MR} + 0.82(\pm 0.27)I_3 + 7.88(\pm 1.02)$	36	0.588	0.511	0.318
	9f	$\log 1/\mathrm{IC}_{50} = 0.55(\pm 0.48)I_1 + 0.67(\pm 0.32)I_3 + 5.36(\pm 0.16)$	36	0.388	0.309	0.387

OEN = original equation number; NSQ = number of QSAR models obtained by the use of less number of descriptors than that of the original QSAR.

[a] Eq. (7b) can be considered, but the statistics of Eq. (7) are far better than that of Eq. (7b).

Table 9
QSAR models of the training sets (original sets−test sets) with their statistical parameters, and the comparison between the observed and predicted biological activities of the test set compounds

OEN	QSAR of the training sets (original sets−test sets) used to predict the biological activities of the test set compounds	Biological activities of the test set compounds (randomly selected)			
		Compd. no.	log 1/IC$_{50}$		
			Obsd.	Pred.	Δ
(1)	$\log 1/\mathrm{IC}_{50} = 0.51(\pm 0.44)\pi_{\mathrm{R}} + 5.12(\pm 0.83)$; $n = 4$, $r^2 = 0.925$, $s = 0.184$, $q^2 = 0.737$, $Q = 5.228$, $F_{1,2} = 24.667$	**1–3**	6.16	5.92	0.24
		1–5	6.55	6.13	0.42
(2)	$\log 1/\mathrm{IC}_{50} = 0.50(\pm 0.42)\pi_{\mathrm{R}} + 4.75(\pm 0.85)$; $n = 4$, $r^2 = 0.930$, $s = 0.159$, $q^2 = 0.569$, $Q = 6.063$, $F_{1,2} = 26.571$	**1–2**	4.79	5.25	−0.46
		1–3	5.76	5.52	0.24
(3)	$\log 1/\mathrm{IC}_{50} = -0.17(\pm 0.10)\pi_X + 0.40(\pm 0.23)\mathrm{MR}_X + 5.69(\pm 0.13)$; $n = 7$, $r^2 = 0.896$, $s = 0.063$, $q^2 = 0.748$, $Q = 15.016$, $F_{2,4} = 17.231$	**2–11**	5.86	6.06	−0.20
		2–14	6.12	6.04	0.08
		2–17	5.95	6.14	−0.19
(4)	$\log 1/\mathrm{IC}_{50} = 0.17(\pm 0.07)\pi_X + 0.12(\pm 0.05)E_{\mathrm{S}-X} + 5.71(\pm 0.10)$; $n = 7$, $r^2 = 0.941$, $s = 0.040$, $q^2 = 0.651$, $Q = 24.250$, $F_{2,4} = 31.898$	**3–19**	5.55	5.70	−0.15
		3–25	5.29	5.53	−0.24
		3–28	5.82	5.90	−0.08
(5)	$\log 1/\mathrm{IC}_{50} = -0.84(\pm 0.57)\sigma_X^+ + 5.73(\pm 0.27)$; $n = 7$, $r^2 = 0.738$, $s = 0.244$, $q^2 = 0.547$, $Q = 3.521$, $F_{1,5} = 14.084$	**3–18**	5.87	5.73	0.14
		3–27	6.14	5.98	0.16
(6)	$\log 1/\mathrm{IC}_{50} = 6.37(\pm 3.18)\pi_X - 8.88(\pm 4.48)\pi_X^2 + 1.48(\pm 1.24)\sigma_X^+ + 5.28(\pm 0.19)$; $n = 9$, $r^2 = 0.857$, $s = 0.124$, $q^2 = 0.503$, $Q = 7.468$, $F_{3,5} = 9.988$, optimum $\pi_X = 0.36(0.34 - 0.39)$	**4–33**	5.72	5.60	0.12
		4–36	5.93	5.89	0.04
		4–39	5.64	5.48	0.16
(7)	$\log 1/\mathrm{IC}_{50} = -0.83(\pm 0.57)\pi_X + 1.43(\pm 0.65)\sigma_X^+ + 6.13(\pm 0.31)$; $n = 10$, $r^2 = 0.857$, $s = 0.196$, $q^2 = 0.593$, $Q = 4.725$, $F_{2,7} = 20.976$	**4–29**	6.19	6.13	0.06
		4–37	5.87	5.91	−0.04
		4–41	4.79	4.76	0.03
		4–43	4.91	4.76	0.15
(8)	$\log 1/\mathrm{IC}_{50} = 0.15(\pm 0.08)\log P - 0.30(\pm 0.19)I_3 - 0.40(\pm 0.17)I_4 + 5.19(\pm 0.41)$; $n = 28$, $r^2 = 0.667$, $s = 0.189$, $q^2 = 0.549$, $Q = 4.323$, $F_{3,24} = 16.024$	**1–1**	5.53	5.45	0.08
		1–4	6.33	6.17	0.16
		1–6	6.53	6.38	0.15
		2–12	6.10	5.86	0.24
		2–15	6.12	5.69	0.43
		3–18	5.73	5.57	0.16
		3–22	5.86	5.75	0.11
		3–27	5.70	5.65	0.05
		4–32	5.61	5.56	0.05
(9)	$\log 1/\mathrm{IC}_{50} = -0.24(\pm 0.10)\mathrm{MR} + 0.68(\pm 0.46)I_1 + 0.94(\pm 0.31)I_3 + 8.29(\pm 1.21)$; $n = 27$, $r^2 = 0.696$, $s = 0.301$, $q^2 = 0.514$, $Q = 2.771$, $F_{3,23} = 17.553$	**1–5**	5.70	5.81	−0.11
		2–7	5.19	5.27	−0.08
		2–9	5.21	5.15	0.06
		2–12	5.13	5.12	0.01
		2–17	5.12	4.82	0.30
		3–20	5.97	6.09	−0.12
		3–27	6.14	6.10	0.04
		4–35	5.80	5.91	−0.11
		4–36	5.71	5.99	−0.28

OEN = original equation number.

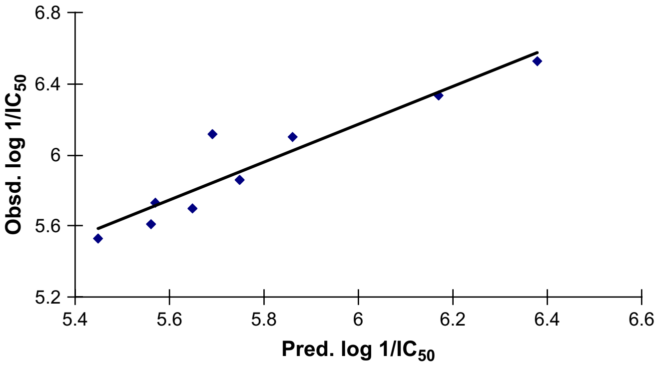


Fig. 10. Plot of observed versus predicted log $1/IC_{50}$ for nine compounds in test set [Eq. (8), Table 9].

q^2 values. This is a widely used technique to ensure the robustness of a QSAR model. The statistical data of r^2 and q^2 for five runs have been listed in Table 7 [Eqs. (1)–(9)]. The poor values of r^2 and q^2 in the Y-randomization test ensure the robustness of QSAR models [31,39–41].

- *Lack of overfitting*: A model overfits if it includes more descriptors than required. The lack of overfitting for all the QSAR models [Eqs. (1)–(9)] was confirmed by using the following conditions.
  - Number of data points/number of descriptors ≥ 4.
  - High values of Q.
  - All the QSAR models [except those with only one descriptor i.e. Eqs. (1), (2), and (5)] were checked for their correlation with a fewer numbers of descriptors than that of the original and listed in Table 8 [42].
  - Y-randomization test (Table 7) suggests that the high r^2 values of the QSAR models [Eqs. (1)–(9)] are not due to the chance correlation or overfitting [43].

4.2. External validation

The original data sets of the QSAR models [Eqs. (1)–(9)] were divided into training ($n = 70$–75%) and test ($n = 25$–30%) sets in a random manner. QSAR models for these training sets were generated by using the same descriptors as those of their respective original QSARs [Eqs. (1)–(9)] and validated from their statistics (acceptance criteria: $r^2 > 0.6$ and $q^2 > 0.5$). The validated QSAR models of these training sets were then used to predict the biological activities of their respective test set compounds. A very good agreement between the predicted and experimental (observed) biological activities of the test set compounds (see Table 9) has validated all the QSAR models [Eqs. (1)–(9)]. It is important to note here that the QSAR models [Eqs. (8) and (9)] have reasonably large size of test sets (each has nine compounds). Thus, the validation of these QSAR models [Eqs. (8) and (9)] can also be demonstrated by the comparison between observed and predicted log $1/IC_{50}$ of their respective nine test set compounds (Table 9) as shown in Figs. 10 and 11. As it can be seen in Figs. 10 and 11, the models exhibit remarkably good predictive abilities ($r^2 = 0.883$ and 0.916).

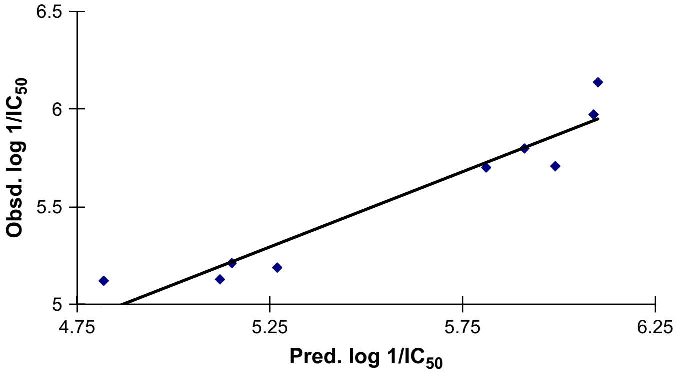


Fig. 11. Plot of observed versus predicted log $1/IC_{50}$ for nine compounds in test set [Eq. (9), Table 9].

5. New molecule prediction

QSAR Eq. (6) is a parabolic correlation in terms of the hydrophobic parameter of the X-substituents where the optimum hydrophobicity of the X-substituents is well defined, whilst activity is also linearly related to σ_X^+ ($\sigma^+$of X-substituents). We believe that this equation may prove to be an adequate predictive model that can help to provide guidance in design/synthesis and subsequently yield very specific organotin compounds **4** that have high kill activities against *Ae. aegypti* mosquito larvae. On the basis of this model, that is, by boosting the electron withdrawing nature of X-substituents and preserving $\pi_X \approx 0.36$ for the compound series **4**, nine compounds **4a**–**4i** are suggested as potential synthetic targets. For the validation of the predictive biological activities of these suggested compounds **4a**–**4i**, we calculated their predicted log $1/IC_{50}$ from both of the equations i.e. the original QSAR Eq. (6) and the QSAR of their training set. A very good agreement between the predicted biological activities obtained from both of the equations (Table 10) has validated the predictive biological activities of these suggested compounds **4a**–**4i**. It can also be demonstrated by Fig. 12, which exhibits an excellent predictive ability ($r^2 = 0.999$). The predictive capacity of the QSAR Eq. (6) is further judged from their high predictive R^2 ($R^2_{\text{Pred}} = 0.864$) value, which was calculated by Eq. (10):

$$R^2_{\text{Pred}} = 1 - \frac{\sum\left(Y_{\text{Pred(Test)}} - Y_{\text{Test}}\right)^2}{\sum\left(Y_{\text{Test}} - \overline{Y}_{\text{Training}}\right)^2} \tag{10}$$

In the above Eq. (10), $Y_{\text{Pred(Test)}}$ and $Y_{\text{(Test)}}$ are the predicted and observed activities, respectively, of the test set compounds and $\overline{Y}_{\text{Training}}$ is the mean activity of the training set compounds.

It is interesting to note here that these compounds **4a**–**4i** are also well predicted by Eq. (7), which suggests that these molecules may also have the high kill activities against *An. stephensi* mosquito larvae. For the validation of the predictive biological activities of these compounds **4a**–**4i**, we also calculated their predicted log $1/IC_{50}$ from both of the equations i.e. the original QSAR Eq. (7) and the QSAR of their training set. Again a very good agreement between the predicted biological activities obtained from both of the equations (Table 10) has

Table 10
Predicted log $1/IC_{50}$ of tris-(*para*-substituted phenyl)tins **4** against *Aedes aegypti* and *Anopheles stephensi* mosquito larvae obtained from Eqs. (6) and (7) as well as from the equations of their training sets along with their physicochemical parameters

No.	X	Y	log $1/IC_{50}$ (Pred.)		log $1/IC_{50}$ (Pred.)		π_X	σ_X^+
			Eq. (6) [a]	Eq. (6) [b]	Eq. (7) [a]	Eq. (7) [b]		
4a	C≡CH	Cl	6.70	6.67	6.04	6.06	0.40	0.18
4b	C≡CH	OH	6.70	6.67	6.04	6.06	0.40	0.18
4c	C≡CH	OAc	6.70	6.67	6.04	6.06	0.40	0.18
4d	$COOC_2H_5$	Cl	6.97	6.92	6.36	6.40	0.51	0.48
4e	$COOC_2H_5$	OH	6.97	6.92	6.36	6.40	0.51	0.48
4f	$COOC_2H_5$	OAc	6.97	6.92	6.36	6.40	0.51	0.48
4g	4-Pyridinyl	Cl	8.11	8.06	7.34	7.43	0.46	1.17
4h	4-Pyridinyl	OH	8.11	8.06	7.34	7.43	0.46	1.17
4i	4-Pyridinyl	OAc	8.11	8.06	7.34	7.43	0.46	1.17

[a] Predicted log $1/IC_{50}$ from the original Eqs. (6) and (7).

[b] Predicted log $1/IC_{50}$ from the training set equations of the original data sets of Eqs. (6) and (7).

validated the predictive biological activities of these compounds **4a–4i**. It can also be demonstrated by Fig. 13, which exhibits an excellent predictive abilities ($r^2 = 0.999$). The predictive capacity of the QSAR Eq. (7) is further judged from their high predictive R^2 ($R^2_{Pred} = 0.981$) value, which was calculated by Eq. (10).

The predicted log $1/IC_{50}$ of these compounds **4a–4i** against *Ae. aegypti* and *An. stephensi* mosquito larvae obtained from Eqs. (6) and (7) as well as from the equations of their training sets along with their physicochemical parameters are given in Table 10.

6. Conclusion

In this paper, we derived seven equations from four different compound series of organotins **1–4** for their larvicidal activities against the second instar stage of the *Ae. aegypti* and *An. stephensi* mosquito larvae. Finally, we combined the data and developed two more equations to understand the chemical–biological interactions. Our QSAR results suggest that the two most important determinants for the toxicity are the hydrophobic (π) and Hammett electronic (σ^+) parameters of the substituents, and the kill mechanism is different for these two species of mosquito larvae.

Out of nine QSAR models, seven contain a correlation between toxicity and hydrophobicity. A positive linear correlation is found in four equations [Eqs. (1), (2), (4), and (8)]. The coefficient with the hydrophobic parameter varies considerably, from a low value of 0.15 [Eq. (8)] to the high value of 0.58 [Eq. (2)]. These data suggest that toxicity might be improved by increasing substituent/compound hydrophobicity. A negative linear correlation is found in two equations [Eqs. (3) and (7)], the coefficients are −0.21 [Eq. (3)] and −0.83 [Eq. (7)]. Less hydrophobic congeners in these compound families might display enhanced toxicity. Parabolic correlation with hydrophobic parameter of the substituents is found in one equation [Eq. (6)], which reflects situation where toxicity increases with increasing hydrophobicity of the substituents up to an optimal value and then decreases. This is an encouraging example, where the optimal value of hydrophobic parameter of the substituents is well defined. We believe that this equation may prove to be adequate predictive model to narrow the synthetic challenges in order to yield very specific compounds **4** that may have high larvicidal activities against *Ae.*

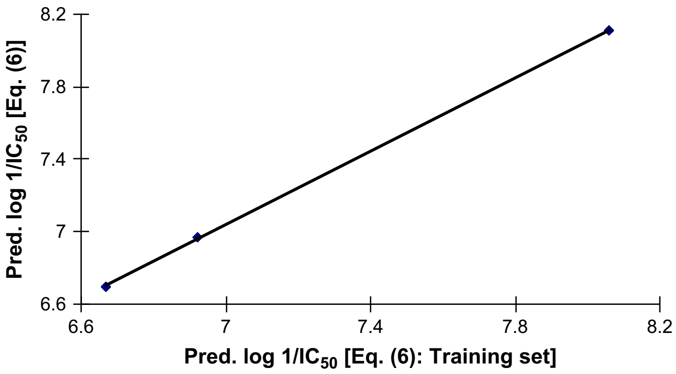


Fig. 12. Plot of the predicted biological activities (log $1/IC_{50}$) of suggested compounds **4a–4i** against *Aedes aegypti* obtained from the original QSAR Eq. (6) and the equation of their training set [Note: only three points instead of nine points are shown because there are only three different sets of values for log $1/IC_{50}$ i.e. each set of three compounds have the identical values of log $1/IC_{50}$ (see Table 10)].

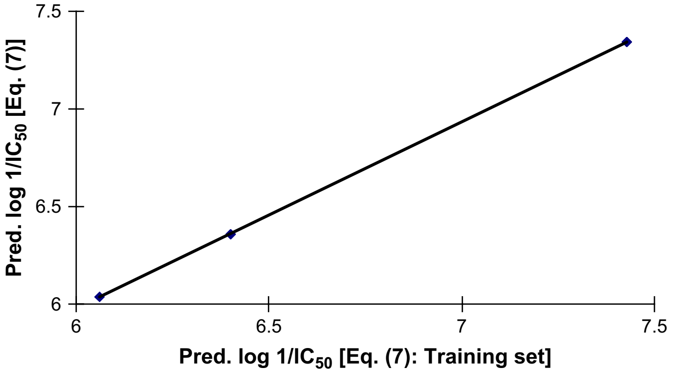


Fig. 13. Plot of the predicted biological activities (log $1/IC_{50}$) of suggested compounds **4a–4i** against *Anopheles stephensi* obtained from the original QSAR Eq. (7) and the equation of their training set [Note: only three points instead of nine points are shown because there are only three different sets of values for log $1/IC_{50}$ i.e. each set of three compounds have the identical values of log $1/IC_{50}$ (see Table 10)].

aegypti mosquito larvae. On the basis of this model, nine compounds **4a–4i** are suggested as potential synthetic target. These compounds are also well predicted by Eq. (7) suggesting that these molecules may also have high larvicidal activities against *An. stephensi* mosquito larvae.

The second important parameter is Hammett electronic parameter (σ^+), which is present in three equations [Eqs. (5)–(7)]. Molar refractivity of the substituent/compound [Eqs. (3) and (9)] and Taft's steric parameter [Eq. (4)] are also important, but they do not seem to play as important a role as hydrophobicity and Hammett electronic parameter (σ^+) of the substituents for the data sets that we have examined. Further development of QSAR studies should not only enlarge the areas of their application, but also increase our understanding towards the mechanisms of chemical–biological interactions. This may assist in the development of new compounds with excellent larvicidal activity.

References

[1] A.A. James, Science 257 (1992) 37–38.
[2] N.G. Das, D. Goswami, B. Rabha, J. Vect. Borne Dis. 44 (2007) 145–148.
[3] A.W.A. Brown, J. Am. Mosq. Control Assoc. 2 (1986) 123–140.
[4] S.E. Robertson, B.P. Hull, O. Tomori, O. Bele, J.W. LeDuc, K. Esteves, J. Am. Med. Assoc. 276 (1996) 1157–1162.
[5] D. Butler, J. Maurice, C. O′Brien, Nature 386 (1997) 535–536.
[6] S.I. Kim, O.K. Shin, C. Song, K.Y. Cho, Y.J. Ahn, Agric. Chem. Biotechnol. 44 (2001) 23–26.
[7] K. Sukumar, M.J. Perich, L.R. Boobar, J. Am. Mosq. Control Assoc. 7 (1991) 210–237.
[8] R. Babu, K. Murugan, Neem Newslett. 15 (1998) 9–11.
[9] M.R. Venketachalam, A. Jebasan, J. Exp. Zool. India 4 (2001) 99–101.
[10] S.C. Dwivedi, K. Karwasara, Ind. J. Entomol. 65 (2003) 335–338.
[11] S.J. Blunden, in: P.J. Craig (Ed.), Organometallic Compounds in the Environment, Wiley, New York, 1986, p. 111.
[12] X. Song, A. Zapata, J. Hoerner, A.C. de Dios, L. Casabianca, G. Eng, Appl. Organomet. Chem. 21 (2007) 545–550.
[13] T.S.B. Baul, K.S. Singh, A. Lycka, A. Linden, X. Song, A. Zapata, G. Eng, Appl. Organomet. Chem. 20 (2006) 788–797.
[14] Q. Duong, X. Song, E. Mitrojorgji, S. Gordon, G. Eng, J. Organomet. Chem. 691 (2006) 1775–1779.
[15] T.S.B. Baul, K.S. Singh, M. Holcapek, R. Jirasko, A. Linden, X. Song, A. Zapata, G. Eng, Appl. Organomet. Chem. 19 (2005) 935–944.
[16] X. Song, Q. Duong, E. Mitrojorgji, A. Zapata, N. Nguyen, D. Strickman, J. Glass, G. Eng, Appl. Organomet. Chem. 18 (2004) 363–368.
[17] G. Eng, X. Song, Q. Duong, D. Strickman, J. Glass, L. May, Appl. Organomet. Chem. 17 (2003) 218–225.
[18] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[19] C. Hansch, A. Leo, in: Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, D.C., 1995, pp. 169–543.
[20] C.D. Selassie, R. Garg, S. Kapur, A. Kurup, R.P. Verma, S.B. Mekapati, C. Hansch, Chem. Rev. 102 (2002) 2585–2605.
[21] R.P. Verma, A. Kurup, S.B. Mekapati, C. Hansch, Bioorg. Med. Chem. 13 (2005) 933–948.
[22] (a) C. Hansch, A. Leo, S.B. Mekapati, A. Kurup, Bioorg. Med. Chem. 12 (2004) 3391–3400;
(b) R.P. Verma, C. Hansch, C.D. Selassie, J. Comput. Aided Mol. Des. 21 (2007) 3–22.
[23] C.D. Selassie, S.B. Mekapati, R.P. Verma, Curr. Top. Med. Chem. 2 (2002) 1357–1379.
[24] C-QSAR Program, BioByte Corp., 201W. 4th st., Suit 204, Claremont, CA 91711, USA. Available from: www.biobyte.com.
[25] C. Hansch, D. Hoekman, A. Leo, D. Weininger, C.D. Selassie, Chem. Rev. 102 (2002) 783–812.
[26] R.P. Verma, C. Hansch online publication, Nat. Protoc. (2007). doi:10.1038/nprot.2007.125, <http://www.natureprotocols.com/2007/03/05/development_of_qsar_models_usi_1.php> Available from:.
[27] R.D. Cramer III, J.D. Bunce, D.E. Patterson, I.E. Frank, Quant. Struct.-Act. Relat. 7 (1988) 18–25.
[28] L. Pogliani, Chem. Rev. 100 (2000) 3827–3858.
[29] L. Pogliani, J. Phys. Chem. 100 (1996) 18065–18077.
[30] V.K. Agrawal, J. Singh, P.V. Khadikar, C.T. Supuran, Bioorg. Med. Chem. Lett. 16 (2006) 2044–2051.
[31] C.D. Selassie, S. Kapur, R.P. Verma, M. Rosario, J. Med. Chem. 48 (2005) 7234–7242.
[32] R.P. Verma, C. Hansch, Mol. Pharm. 3 (2006) 441–450.
[33] R.P. Verma, C. Hansch, Virology 359 (2007) 152–161.
[34] (a) R.P. Verma, C. Hansch, Bioorg. Med. Chem. 15 (2007) 2223–2268;
(b) R.P. Verma, C. Hansch, J. Pharm. Sci. 97 (2008) 88–110.
[35] S.J. Blunder, P.A. Cusack, R. Hill, The Industrial Uses of Tin Chemicals, Royal Society of Chemistry, London, 1985, 83.
[36] S. Kapur, A. Shusterman, R.P. Verma, C. Hansch, C.D. Selassie, Chemosphere 41 (2000) 1643–1649.
[37] J.F. Contrera, P. MacLaughlin, L.H. Hall, L.B. Kier, Curr. Drug Discov. Technol. 2 (2005) 55–67.
[38] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.
[39] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci. 22 (2003) 69–77.
[40] S. Wold, L. Eriksson, in: H. van de Waterbeemd (Ed.), Statistical Validation of QSAR Results, Chemometrics Methods in Molecular Design, Wiley-VCH, Weinheim, 1995, pp. 309–318.
[41] (a) M.F. Melzig, G.D. Tran, K. Henke, C.D. Selassie, R.P. Verma, Pharmazie 60 (2005) 869–873;
(b) C. Hansch, R.P. Verma, Chem. Med. Chem. 2 (2007) 1807–1813.
[42] C. Hansch, R.P. Verma, Mol. Pharm. 5 (2008) 151–161.
[43] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, A. Tropsha, J. Chem. Inf. Model. 46 (2006) 1984–1995.