International Journal of Civil Engineering and Environmental Science. Volume 2, Issue 2, (2014) 62-70
http://www.cayley-nielson.com/ijcees/

QSAR Study on Pheromone of the Turnip Moth Agrotis Segetum

Mao Li
School of Science, Nankai University
E-mail: *maoli@nankai.edu.cn*

Abstract

Quantitative information on bioactive biological pheromone analogs from the turnip moth *Agrotis segetum* was studied, and the best prediction model was determined. The data set contained 45 organic molecules, of which 35 chemical compounds were selected as test set, and the other 10 were selected as the training set to build a quantitative structure–activity relationship model. For each analog, 150 molecular parameters were calculated, and multiple linear regression analysis was used to build the best model (used in the training and test sets, with correlation coefficients of R^2= 0.898 and 0.869, respectively). The linear relationship between biological activity and log*P* was also tested (R^2 = 0.245). Our results can serve as a reference for the quantitative prediction of pheromone activity and for the design of a new pesticide.

Keywords: Agrotis segetum pheromone, molecular parameter, multiple linear regression, QSAR.

1. Introduction

Communication between individuals comes in a variety of ways, such as by sound, sight, scent, and so on. The oldest way to communicate is through chemical signal substances released by individuals. from simple single-celled organisms (such as bacteria, algae, and fungi) to the highly developed humans, chemical signals are essential and apparent. Nobel Prize awardee A. Butenandt [1] first extracted a sexually stimulating compound from the gonads of a female aphid, which is stimulated by mating and makes them attract each other. He ultimately determined the existence of sexually stimulating compounds through his continuous research in 1959. In the same year, Karlson and Luscher [2] first proposed the use of term pheromone to define the compounds and method of exchanging information between individuals through chemical signals, and established a new field of study. Study of insect pheromones is widely practiced as it does not only elucidate the chemical structure, biosynthetic pathway, molecular basis of pheromones, but it also supports the extensive research involving receptor structure and biological function.

Pheromones are sexual compounds of higher forms of living organism that are used to recognize each other. These substances can make the female and the male attract and mate with one another. Generally, pheromones are released by passive females to generate excitement and lure males. However, some species males also release pheromones. A. Butenandt [3] isolated bombyxin alcohol from the female silkworm, and determined that it is a trans-10,cis-12,16-carbon diene-1-alcohol, and also researched Lepidoptera. Pheromones contain straight-chain alcohols or aceticyl with 12 to 16 carbon atoms and have one or two double bonds. In addition, chemical structures of some pheromones of *Coleoptera* and *Orthoptera* have been identified, but only show slight differences. Mammalians also have pheromones,

ISSN 2333-5416
December, 2014

which are as studied in the fields of biology and chemistry. Many examples have been shown such as the relationship of spousal behaviors with a number of pheromones. Pheromone chemical structures are identical among heterologous animals, and their similarities have been studied for pest control.

Quantitative structure–activity relationship (QSAR) is used to describe the relationship between molecular structure and biological activity. The basic assumption is that the molecular structure helps determine physical, chemical, and biological nature of the compound, which then determine the its biological activity. QSAR is a reliable and time- and labor-saving method that can be used in study of pheromones, which provides a means of predicting the functions and activities of chemical signaling compounds using available information on their molecular structures. These have important theoretical and practical roles in the in-depth study of the relationship between these biological signaling compounds and their biological activities. The purpose of this study is to create a new linear QSAR model to predict the biological activity of pheromones. With a reasonable choice of physical, and chemical, and molecular structure parameters. A better model with excellent reliability and predictability will be proposed using multiple linear regression method.

2. Date Sets and Methods

2.1 Date sets

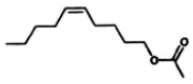
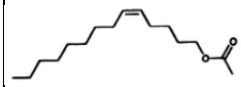
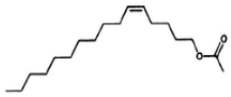
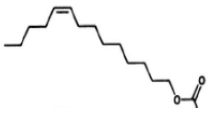
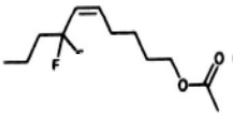
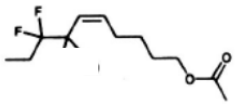
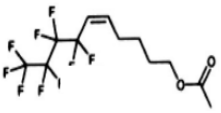
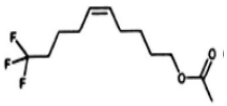
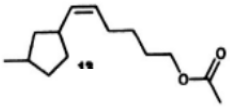
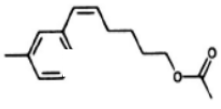
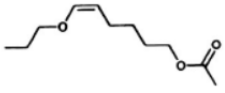
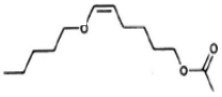
All the data regarding pheromone activities used in this article were from the literature. Compounds 1 to 7 were from Liljefors et al. [4], compounds 8 to 12 were from Wenqui et al.[5], compounds 13 and 14 were from Johnson et al.[6], compounds 15 to 18 were from Gustavsson et al.[7], compounds 19 to 29 were from Johnson et al.[8-10], compounds 30 to 34 were from Bengtsson et al.[11], compounds 35 and 36 were from an unpublished study (B. Hansson, Lund University), compounds 37 to 39 were from Ge Siwen et al. [12], compounds 40 to 43 were from Johnson et al.[9-10], and compounds 44 and 45 were from Johnson et al. [13].

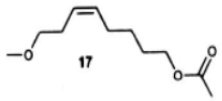
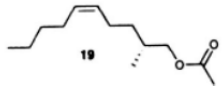
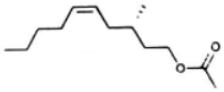
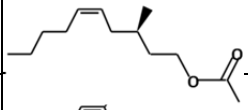
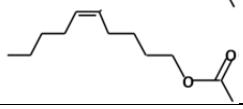
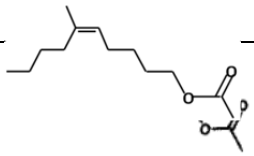
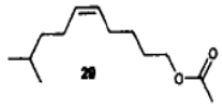
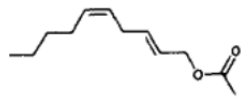
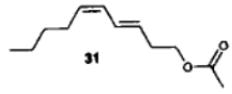
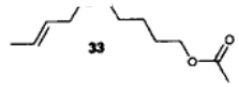
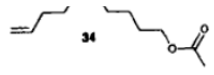
In the modeling process, 75% of the sample data were included in the training set and used to create a training model, while the remaining data were included in the test set. The best prediction results were obtained by repeated training and testing, and statistical verification. The chemical structures and biological activity data of the training and validation sets are shown in Tables 1 and 2, respectively.

2.2 Parameter calculation

All compounds in the present study were first converted into 2D structures using Chemdraw from Chemoffice software. These were then converted into SMILES format and entered into a molecular descriptor calculation software called The Dragon (http://www.talete.mi.it/) to simulate the compound molecules, and obtain the most commonly used 150 molecular Radial Distribution Function (RDF) series of the structural parameters. In accordance with the requirements of the prediction method, all the structural parameters 0 and 999 were deleted, and 158 parameters were streamlined. These parameters and the biological activity of the compounds were associated using multiple linear regression method. RDF parameters include: (1) RDF***u index: (Radial Distribution Function - *** / unweighed); (2) RDF***m index: (Radial Distribution Function*** / weighed by atomic masses); (3) RDF***v index: (Radial Distribution Function - *** / weighed by atomic van der Waals volumes); (4) RDF015e index: (Radial Distribution Function - *** / weighed by atomic Sanderson electronegativities);and (5) RDF***p index: (Radial Distribution Function - *** / weighed by atomic polarizabilities) and other descriptor.

Table 1. Chemical structures, octanol/water partition coefficients and biological activities of the pheromones in training set.

No	STRUCTURE	SMILES	log*P*	Log(Exp)	Log(Pre)
1		CCCC\C=C/CCCCOC(C)=O	3.38	7.0	6.33
3		CCCCCCCC\C=C/CCCCOC(C)=O	5.05	2.5	3.42
4		CCCCCCCCCC\C=C/CCCCOC(C)=O	5.89	4.8	4.36
6		CCCC\C=C/CCCCCCCCOC(C)=C	5.05	2.0	2.36
8		CCCC\C=C/C(F)(F)CCCOC(C)=O	3.26	4.9	4.98
9		CCCC(F)(F)\C=C/CCCCOC(C)=O	3.26	5.1	4.86
10		CCC(F)(F)C(F)(F)\C=C/CCCCOC(C)=O	3.44	5.1	5.11
11		CC(=O)OCCCC\C=C/C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	4.37	2.0	1.88
12		CC(=O)OCCCC\C=C/CCCC(F)(F)F	3.68	4.9	4.36
13		CC1CCC(C1)\C=C/CCCCOC(C)=O	3.54	4.9	3.89
14		CC(=O)OCCCC\C=C/C1=CC=CC(=C1)C	3.81	4.0	4.59
15		CCCO\C=C/CCCCOC(C)=O	2.15	4.8	5.41
16		CCCCCO\C=C/CCCCOC(C)=O	2.98	5.0	5.00

17		COCC\C=C/CCCCOC(C)=O	1.71	6.2	5.94
18		COCCCC\C=C/CCCCOC(C)=O	2.54	4.3	3.93
19		CCCC\C=C/CCC(C)COC(C)=O	3.78	4.8	5.46
21		CCCC\C=C/CC(C)CCOC(C)=O	3.71	4.6	4.99
22		CCCC\C=C/CC(C)CCOC(C)=O	3.71	5.4	4.99
23		CCCC\C=C(\C)CCCCOC(C)=O	3.56	4.1	4.02
24		CCCC\C(C)=C/CCCCOC(C)=O	3.56	6.0	5.64
26		CCCC(C)\C=C/CCCCOC(C)=O	3.71	4.1	4.14
27		CCC(C)C\C=C/CCCCOC(C)=O	3.71	5.1	5.24
29		CC(C)CC\C=C/CCCCOC(C)=O	3.71	6.0	6.06
30		CCCC\C=C/C/C=C/COC(C)=O	3.20	5.0	5.18
31		CCCC\C=C/C=C/CCOC(C)=O	3.06	5.8	5.30
32		CC\C=C\C=C/CCCCOC(C)=O	3.06	4.1	4.76
33		C\C=C\C\C=C/CCCCOC(C)=O	3.06	6.2	6.15
34		CC(=O)OCCCC\C=C/CCC=C	3.11	6.0	5.70

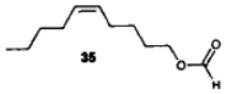
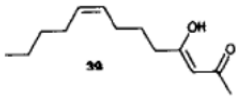
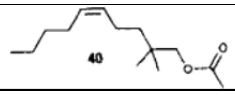
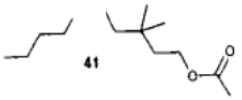
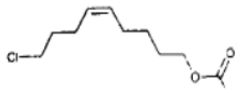
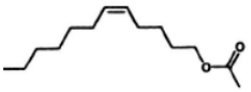
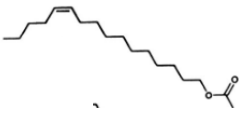
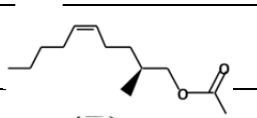
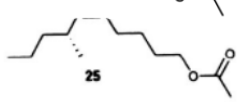
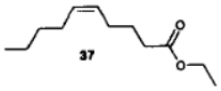
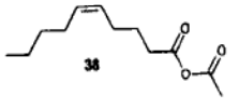
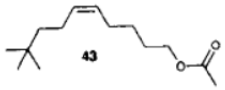
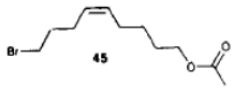
35		CCCC\C=C/CCCCOC=O	3.12	4.6	4.50
36		CCCC\C=C/CCCCOC(=O)CC	4.04	5.6	4.73
39		CCCC\C=C/CCC\C(O)=C\C(C)=O	3.23	5.4	5.83
40		CCCC\C=C/CCC(C)(C)COC(C)=O	4.32	4.2	4.80
41		CCCC\C=C/CC(C)(C)CCOC(C)=O	4.18	3.8	4.03
42		CCC(C)(C)C\C=C/CCCCOC(C)=O	4.18	5.0	4.55
44		CC(=O)OCCCC\C=C/CCCCl	2.99	5.3	6.06

Table 2. Chemical structure, biological activity and octanol/water partition coefficient of the pheromone in test set.

No	**STRUCTURE**	**SMILES**	**log*P***	**Log(Exp)**	**Log(Pre)**
2		CCCCCC\C=C/CCCCOC(C)=C	4.22	4.2	4.15
5		CCCC\C=C/CCCCCCOC(C)=O	4.22	4.5	4.95
7		CCCC\C=C/CCCCCCCCCCOC(C)=O	5.89	4.0	5.19
20		CCCC\C=C/CCC(C)COC(C)=O	3.78	4.6	5.46
25		CCCC(C)\C=C/CCCCOC(C)=O	3.71	5.3	4.14
28		CCC(C)C\C=C/CCCCOC(C)=O	3.71	5.1	5.24

37		CCCC\C=C/CCCC(=O)OCC	3.55	3.5	5.58
38		CCCC\C=C/CCCC(=O)OC(C)=O	2.92	4.7	5.92
43		CC(=O)OCCCC\C=C/CCC(C)(C)C	4.18	5.1	4.73
45		CC(=O)OCCCC\C=C/CCCBr	3.11	5.6	6.13

2.3 Mathematical tools

This study was based on the molecular parameters of multiple linear regression (MLR) analysis. MLR is based on statistical analysis and is used to calculate regression equation. It also the earliest computational modeling method used to study QSAR. The basic assumptions in in MLR are based on the changes in the molecular structures and biological activities, which are related to the physical and chemical parameters. The MLR equation is as follows:

$$Y = \beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k \quad (1)$$

where $\beta_1,\beta_2,...,\beta_k$ are the regression coefficients. β_i is the average change in the dependent variable Y arising from the variation of X_k, (i = 1, 2, …k), in the case where other independent variable remain unchanged.

3. Results

3.1 Results of multiple linear regression analysis

In this paper, a stepwise method was chosen and used to calculate and establish the predictive model, which had a better statistical result (R = 0.948) and a better predictability (R^2 = 0.898). The linear regression relationship was significant between the structural parameters of the compound molecules and their corresponding biological activities.

A mathematical model based on six parameters (Table.3) was obtained when stepwise regression analysis was conducted. The model can be defined by the equation:

$$Y = 6.756-0.559RDF145m-0.176RDF060u+0.151RDF080e-0.919RDF125m+2.821RDF150m-0.102RDF090u \quad (2)$$

$$n = 35, R^2 = 0.898, S = 0.528, F = 19.695, P < 0.001.$$

where n is the number of samples, $R^2$is the regression coefficient, S is the standard deviation, and F is Fischer test value of the model (P<0.001), which is statistically significant.

Table 3 shows the parameters that include RDF145m[1272], RDF060u[1225], RDF080e[1319], RDF125m[1268], RDF150m[1273], and RDF090u[1231].

3.2 Detection of the model

Figure.1 shows the comparison of experimental data of the stepwise and predicted values, where the triangle and square pertain to the training set and the test set molecules respectively. The results show that

the predicted and experimental values are in good agreement. The abscissa refers to the test values, and the ordinate refers to the predicted values.

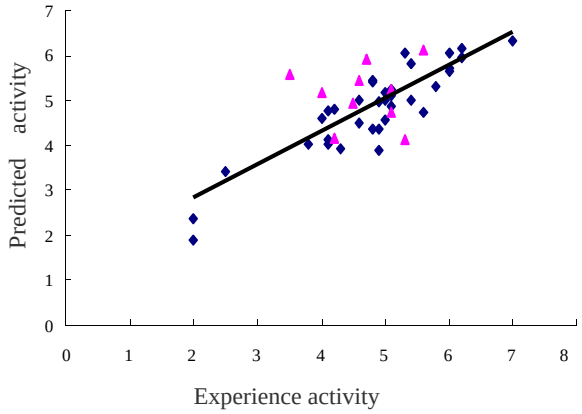


Fig.1. Experimental and predicted values of the stepwise method.

3.3 Relationship between the biological activity and octanol/water partition coefficient

In this study, each octanol/water partition coefficient (log*P*) was tested for a linear relationship with the biological activity of the pheromones. The biological activity and log*P* relations were analyzed using statistics, and the results show that there was no significant linear relationship between the two (R = 0.49 and R^2 = 0.245). The results of statistical analysis were shown in Figure 2.

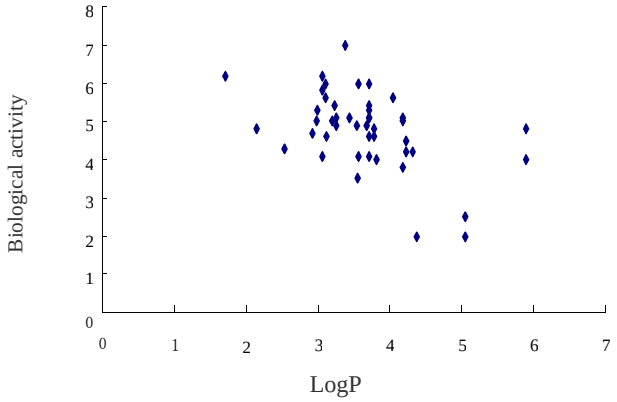


Fig.2. Relationship between biological activity and logP.

Table 3. Definition of model descriptors using the stepwise method.

Descriptors	Definition	UC
RDF145m[1272]	Radial Distribution Function - 14.5 / weighed by atomic masses	-0.559
RDF080e[1319]	Radial Distribution Function - 8.0 / weighed by atomic Sanderson electroneg activities	0.151
RDF060u[1225]	Radial Distribution Function - 6.0 / unweighed	-0.176
RDF080e[1319]	Radial Distribution Function - 8.0 / weighed by atomic Sanderson electroneg activities	0.151

RDF125m[1268]	Radial Distribution Function - 12.5 / weighed by atomic masses	-0.919
RDF150m[1273]	Radial Distribution Function - 15.0 / weighed by atomic masses	2.821
RDF090u[1231]	Radial Distribution Function - 9.0 / unweighed	-0.102

*UC: Unstandardized Coefficients

4. Disscussion

Pheromone binding proteins (PBPs) have a double role of carrying and deactivating in the insect antennae. PBPs dissolve and transport fat-soluble pheromones through the sensor's hydrophilic lymph to reach the dendritic membrane, and then deactivate the pheromones. A reduced PBP first combines with the pheromone and is oxidized once the pheromone and receptor membrane combine. The oxidized PBP then combines and deactivates odor molecules. Therefore, reduced PBP may be used as a pheromone carrier, providing a binding ligand for the receptors of the dendritic membranes. Oxidized pheromone-PBP complexes do not stimulate the receptor cells. Receptor-mediated pheromone-PBP complexes may be the first step in the deactivation. The three main functions of PBP are as follows: First, it transports pheromone molecules in the lymph via the sensor micropores; second, it participates in the removal of the pheromone metabolites; third and last, it complexes with pheromone molecules to play a role in the G-protein coupled receptors of the dendritic membranes to activate the signaling pathway. The PBP complexes and pheromones combine with the receptors in the dendritic membrane, and activate the receptor-mediated G-protein, which sequentially activates the key enzyme in the second messenger cascade reaction. Adenylate cyclase catalyzes the conversion of ATP to cAMP. Phospholipase C hydrolyzes the membrane phosphatidylinositol, thereby releasing 1,4,5-inositol triphosphate) and diacylglycerol. Ion channels in plasma membranes are activated as the media concentration rapidly increase, and nerve impulses and sensor potential emerge.

A recent study of involving PBP [17,18] could provide another explanation, which estimates their activities 3D QSAR model. PBP complexes pheromone molecules and transports them to the receptors. Recent studies showed that PBP combined with the natural pheromone component have a higher affinity than the analog [16]. Even having the same concentration as with the analog, the pheromone component may have a significantly higher concentration in the receptors. Experimental receptor activity is based on a numbers contest for the same receptor, and because the relative quantity is not certain, the experimental receptor activity may overestimate the high-affinity PBP ligand. If the PBP ligand binding have better affinity data, this problem may be solved.

QSAR is based on the traditional structure–activity relationship, and combines the physical, chemical, and mathematical methods. The history of the theory can be traced back to Crum-Brown's equation in 1868. The equation states that the physiological activity of compounds could be expressed using the function of the chemical structure. However it did not establish a clear functional model. The earliest implemented QSAR method was the Hansch equation. The Hansch equation grew out of the Hamiltonian equation and improved Taft equation. Hamiltonian equation is an empirical equation which was used in calculating the dissociation constant of a substituted benzoic acid. The equation was also used in establishing a linear relationship between the logarithm of dissociation constant of a substituted benzoic acid and the electrical parameters of substituents. Taft equation is an improvement if the Hamiltonian equation and it was used in calculating the hydrolysis reaction rate constant of aliphatic esters. The equation was also used in establishing a linear relationship between the logarithm of rate constants and electrical parameters, and the three-dimensional parameter.

As the 2D quantitative analysis could not accurately describe the relationship between the 3D molecular structure and its physiological activity, people began to explore the feasibility using 3D QSAR based on molecular conformation in the 1980s. Crippen [14] studied the 3D QSAR of distance geometry in 1979, while Hopfinger et al. [15] studied the molecular shape analysis method in 1980. Moreover, Cramer et al. [16] studied CoMFA in 1988. CoMFA swept the field of drug design when it was first released, and became the most widely used method in drug design that is based on QSAR. In the 1990s, some new 3D QSAR methods, such as CoMSIA (an improvement of CoMFA) and virtual receptor methods based on the 3D QSAR of distance geometry, appeared. However, CoMFA was still the most widely used QSAR method whatever 2D or 3D descriptor methods become available[19]. The essence of a descriptor method is to collect better information on the chemical compounds and obtain better mathematical models. In this present work, the RDF descriptors were used to describe the chemical information of pheromone analogs, and were fitted to the QSAR study.

References

[1] P. Karlson, A. Butenandt, *Entomology*, **4**, 39 (1959)

[2] P. Karlson, M. Luscher, *Nature*, **183**, 55 (1959).

[3] A. Butenandt, R. Beckmann, D. Stamm, *Chemie,* **324**, 84 (1961)

[4] T. Liljefors, M. Engtsson, B.S. Hansson, *J. Chem. Ecol*., **13**, 2023 (1987).

[5] W.W. Qi, M. Bengtsson, B.S. Hansson, T. Liljefors, C. Löfstedt, G.D. Prestwich, W.C. Sun, M. Svensson, *J. Chem. Ecol*., **19**, 143 (1993).

[6] S. Jönsson, B.S. Hansson, T. Liljefors, Bioorg. Med. Chem., **4,**499 (1996).

[7] A.L. Gustavsson, T. Liljefors, B.S. Hansson, *J. Chem. Ecol*.,**21,** 815 (1995).

[8] S. Jönsson, T. Liljefors, B.S. Hansson. *J. Chem. Ecol*., **17,** 103 (1991a).

[9] S. Jönsson, T. Liljefors, B.S. Hansson,. *J. Chem. Ecol*., **18**, 637 (1992).

[10] S. Jönsson, T. Malmström, T. Liljefors, B.S. Hansson, *J. Chem. Ecol*., **19,** 459 (1993).

[11] M. Bengtsson, T. Liljefors, B.S. Hansson, *Bioorg. Chem.*, **15**, 409 (1987).

[12] A.L. Gustavsson, M. Tuvesson, M.C. Larsson, W.W. Qi, B.S. Hansson, T. Liljefors, *J. Chem. Ecol*., **23**, 2755 (1997).

[13] S. Jönsson, T. Liljefors, B.S. Hansson, *J. Chem. Ecol*., **17,** 1381 (1991b).

[14] G.M. Crippen, *J. Chem. Ecol*., **22**, 988 (1979).

[15] A.J. Hopfinger*, J. Amer. Chem. Soc*., **102,** 7196 (1980).

[16] R.D. Cramer, D.E. Patterson, J.D. Bunce, *J. Amer. Chem. Soc*, 110, 5959 (1988).

[17] G.D. Prestwich. G. Du. S.L. Forest, *Chemical Senses,* **20**, 461 (1995).

[18] G. Schiavo, Q.M. Gu, G.D. Prestwich, T.H.Söllner, J.E. Rothman, Proc. National Acad. Sci. USA, **93**, 13327 (1996).

[19] U.Norinder, A.L. Gustavsson, T. Liljefors, *J. Chem. Ecol*., **23**, 12 (1997).