Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 18 (2008) 2854–2859
Bioorganic & Medicinal Chemistry Letters

QSAR study of mosquito repellents from terpenoid with a six-member-ring

Zongde Wang,[a] Jie Song,[b,*] Jinzhu Chen,[a] Zhanqian Song,[c] Shibin Shang,[c] Zhikuan Jiang[d] and Zhaojiu Han[d]

[a]*College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China*
[b]*Department of Chemistry and Biochemistry, University of Michigan-Flint, 303E Kearsley Street, Flint, MI 48502, USA*
[c]*Institute of Chemical Industry of Forest Products, Chinese Academy of Forestry, 16 Suo Jin Bei Road, Nanjing 210042, China*
[d]*Military Medical Institute of Nanjing Command, Nanjing 210002, China*

Received 17 February 2008; revised 27 March 2008; accepted 31 March 2008
Available online 8 April 2008

Abstract—A new class of low-toxicity mosquito repellents is synthesized from α- and β-pinene in terpenoid compounds and preliminary biological tests show promising mosquito repellency. Statistical modeling is built using Codessa in order to reveal the quantitative relationship between the structure and the biological activity and to provide the guidance for further synthetic work.

Mosquito repellents are materials used to disrupt the behavior of blood-seeking mosquitos and provide personal protection against mosquito bites. *N*,*N*-Diethyl-3-methylbenzamide or *N*,*N*-diethyl-*m*-toluamide (DEET or DETA), a widely used mosquito repellent, was first developed in 1951.[1,2] Though products containing DEET have been successfully used over the past sixty years, some recent concerns have attracted public attention to the adverse effects of DEET on human health.[3–6] As a result, the design of less toxic alternatives from a natural source remains an active area of research.

Recently, a new class of alternate mosquito repellents, terpenoid compounds, has been synthesized from turpentine oils whose primary contents are α- and β-pinene (Scheme 1). Certain members of this new class have shown repelling capabilities similar to DEET.[7] One advantage of this class of compounds is the lower toxicity of α- and β-pinene derivatives, allowing their use in products that are easily and safely applied to humans (e.g., perfume).[8] Rational modification of chemical structures for improved mosquito repellency requires deeper understanding of the structure–activity relationship for repellents. Although the qualitative or semi-quantitative relationships of structure and repellency[9–11] have been widely discussed in research, only a small amount of quantitative data have been obtained by Suryanarayana et al.[12] and Katritzky et al.[13] regarding amide analogues of DEET. Suryanarayana et al.[12] correlated mosquito repellency of 40 compounds with three descriptors, lipophilicity, vapor pressure, and molecular length; however, correlation coefficients for these descriptors were low. Katritzky et al.[13] applied Codessa Pro software to the same set of repellents used by Suryanarayana et al.,[12] and correlated chemical structures using different descriptors. In this study, a series of terpenoid compounds containing a six-member-ring were synthesized and QSAR modeling using Codessa 2.7.10[14] was applied to repellent activities.

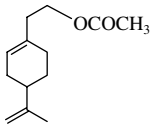
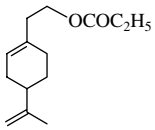
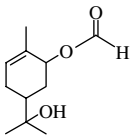
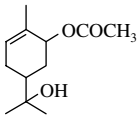
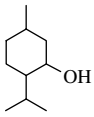
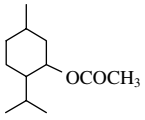
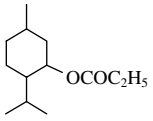
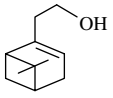
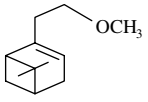
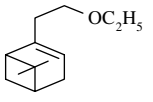
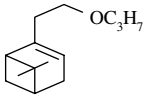
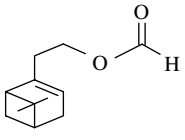
Scheme 1. Structures of α- and β-pinene.

Twenty compounds were synthesized from α- and β-pinene, and are listed in Table 1. All compounds have the base structure of a six-carbon ring with at least

Keywords: QSAR; Mosquito repellents; Terpenoid; Multilinear regression.
* Corresponding author. Tel.: +1 810 762 3110; fax: +1 810 766 6693; e-mail: jiesong@umich.edu

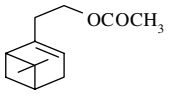
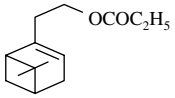
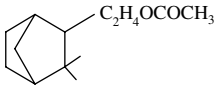
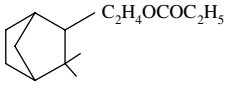
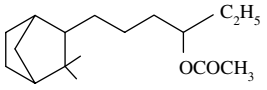
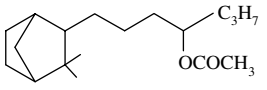
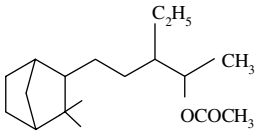
0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.03.091

Table 1. Repellency and descriptors of terpenoid repellents

ID	Structure	CRR [a]	log CRR	log BP [b]	MS [c]	Charge II [d]	log DM [e]
1		58.5	1.7672	2.4522	243.62	0.0070	0.39
2		63.6	1.8035	2.4780	262.02	0.0069	0.37
3		90.0	1.9542	2.4331	210.87	0.0518	0.64
4		81.0	1.9085	2.4451	226.83	0.0483	0.68
5		53.0	1.7243	2.3332	189.47	0.0224	0.24
6		38.7	1.5877	2.3600	214.03	0.0423	0.69
7		40.5	1.6075	2.3913	231.11	0.0412	0.67
8		69.0	1.8388	2.3711	177.63	0.0197	0.22
9		52.1	1.7168	2.3483	198.23	0.0205	0.11
10		62.2	1.7938	2.3851	216.91	0.0206	0.08
11		72.5	1.8603	2.4178	235.19	0.0206	0.04
12		56.1	1.7490	2.4103	204.07	0.0239	0.70

(*continued on next page*)

Table 1 (*continued*)

ID	Structure	CRR [a]	log CRR	log BP [b]	MS [c]	Charge II [d]	log DM [e]
13		60.8	1.7839	2.4294	217.55	0.0219	0.71
14		64.4	1.8089	2.4574	231.99	0.0219	0.69
15		37.8	1.5775	2.2755	158.56	0.0387	0.33
16		53.0	1.7243	2.3969	221.87	0.0236	0.35
17		56.8	1.7543	2.4243	238.26	0.0233	0.32
18		50.2	1.7007	2.4807	286.03	0.0115	0.30
19		51.7	1.7135	2.5015	304.87	0.0115	0.29
20		51.5	1.7118	2.4894	297.11	0.0111	0.32

[a] CRR, corrected repellent ratio at 1.5 h.
[b] BP, boiling point.
[c] MS, molecular surface area.
[d] Charge II, total charge of the other substituent groups on the six-member-ring backbone.
[e] DM, dipole moment.

one O-containing substituent group. There are three kinds of derivatives from α- and/or β-pinene: (1) products from the four-carbon ring opening reaction of the α- or β-pinene[15,16]; (2) the structure of 6,6-dimethylbicyclo[3.1.1]hept-2-ene synthesized from Prins reaction of β-pinene[17,18]; and (3) the structure of 2,2-dimethylbicyclo[2.2.1.]heptane obtained from the isomerization reaction of α-pinene. [7] Synthesis details and the related structure identifications can be found in these references.

In general, descriptors used in QSAR can be categorized as constitutional, topological, geometrical, electrostatic, quantum chemical, and thermodynamic. Previous studies have shown that mosquito repellency is primarily related to the lipophilicity, boiling point (i.e., vapor pressure), and stability. In addition, geometry and charge distribution[12,13,19,20] were found to be closely related to the possible interactions between the repellent and the receptor: either the repellent is bound to the receptors via the electrostatic interactions or the repellent and the receptor have the lock-key fit of many biological processes. Therefore, the relative number of oxygen atoms (OR), HOMO energy, LUMO energy, HOMO–LUMO energy gap, molecular volume, molecular surface, dipole moment, boiling point, charges on positive- and negative-charged ends, and the difference

between these two charges were used as descriptors in this study.

Conformational searches were carried out over all structures and the lowest energy conformers were optimized at the HF/6-31G(d) level using the Gaussian 03 package of programs (version D.01).[21] All quantum chemical descriptors were calculated using Gaussian, although surface area and molecular volume were calculated using AMPAC 8 [22] based upon the optimized geometry. Boiling points, calculated by Advanced Chemistry Development (ACD/Labs) Software, were taken from SciFinder Scholar.[23] All multilinear regressions in this study were done using Codessa 2.7.10.[14]

Bioassay of the terpenoid repellents was previously described in detail.[7] In brief, repellent ratio (RR) was determined by applying test compounds at a dose of 0.16 mg/cm^2 to the depilated abdomen of a white mouse. The mouse was subsequently exposed to 100 female mosquitoes (*Aedes albopictus*), ages 4–5 days old, for 2 min every 30 min with the number of biting mosquitoes being recorded. As an experimental control, ethanol, the repellent solvent, was applied to the depilated abdomen of a white mouse with the number of biting mosquitoes being similarly tested and recorded. The repellent ratio (RR) was calculated using Eq. 1 .

$$\mathrm{RR} = \frac{\mathrm{TN} - \mathrm{TNB}}{\mathrm{TN}} \times 100 \qquad (1)$$

TN is the total number of female mosquitoes used in testing, while TNB is the total number of biting female mosquitoes. The repellent ratio of the control experiment (CERR) was calculated in the same manner as RR. The corrected repellent ratio (CRR) was also calculated in order to evaluate the repellency more accurately (Eq. 2).

$$\mathrm{CRR} = \frac{\mathrm{RR} - \mathrm{CERR}}{100 - \mathrm{CERR}} \times 100 \qquad (2)$$

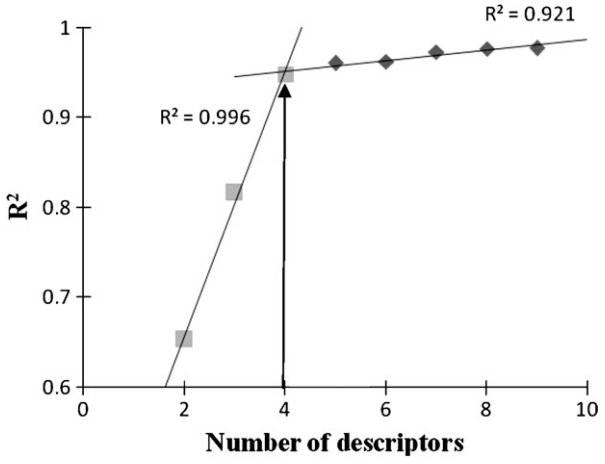


Figure 1. Breaking point rule for determination of the number of the descriptors.

In general, there are many regression approaches available, for example, stepwise regression, principal component analysis, and partial least-square regression.[24] Taking into account the relatively smaller number of samples and descriptors used, in this study, the regression procedure used involves a stepwise reduction in the number of molecular descriptors and *t*-test reflects the significance of the descriptors within the model. Descriptors with high *t* values were accepted, while low *t* values were rejected. The regression calculation was stopped when it came to a 'breaking point', namely when the statistical improvement of the regression model become less significant ($\Delta R^2 < 0.02 - 0.04$). From Figure 1, it is shown that the breaking point occurs at four descriptors, the values of these descriptors being are listed in Table 1.

The statistically best QSAR equation for log CRR data (listed in Table 2) has the following statistical characteristics: $R^2 = 0.95$, $F = 68.8$, $s^2 = 0.0006$. This model includes four descriptors in descending order according to their statistical significance. In Table 2, X and ΔX are the regression coefficients and their standard errors. Table 3 lists values of log CRR obtained using the model described in Table 2, and the graphical representation of these predictions is provided in Figure 2.

Both internal validation and the 'leave-one out' approach were used to validate the obtained model. Internal validation was carried out by dividing the parent data points into three subsets (A ~ C): the first, fourth, seventh, etc., data points going into the first subset (A), the second, fifth, eight, etc., into the second subset (B), and the third, sixth, ninth, etc., into the third subset (C). Two of three subsets, (A and B), (A and C), and (B and C), consist the training set with the remaining subset corresponding to the test set. The correlation equation, derived for each of the training sets using the same descriptors, was used to predict values for the corresponding test sets. Internal validation results can be found in Table 4. Because of the smaller size of test sets, it is not surprising to observe the dispersive predicted R^2. However, the average R^2 (pred) of all the

Table 2. The best four descriptor QSAR model with $R^2 = 0.95$, $F = 68.8$, $s^2 = 0.0006$

Descriptor No.	X	$\pm\Delta X$	*t*-Test	Descriptor [a]
0	−7.0781e+00	5.3312e−01	−3.2769	Intercept
1	4.1474e+00	2.5081e−01	16.5358	log BP
2	−3.3494e−01	3.6260e−02	−9.2373	log DM
3	3.9668e+00	6.4117e−01	6.1868	Charge II
4	−4.9934e−03	3.6356e−04	−3.7350	MS

[a] Descriptors are defined in Table 1.

Table 3. Predicted logCRR

ID	Pred. logCRR	Exp. logCRR	Diff.	ID	Pred. logCRR	Exp. logCRR	Diff.
1	1.7730	1.7672	0.0059	11	1.8432	1.8603	−0.0171
2	1.7942	1.8035	−0.0093	12	1.7597	1.7490	0.0107
3	1.9512	1.9542	−0.0030	13	1.7603	1.7839	−0.0236
4	1.8939	1.9085	−0.0146	14	1.8109	1.8089	0.0020
5	1.6609	1.7243	−0.0634	15	1.6104	1.5775	0.0329
6	1.5778	1.5877	−0.0099	16	1.7310	1.7243	0.0068
7	1.6243	1.6075	0.0168	17	1.7720	1.7543	0.0176
8	1.8730	1.8388	0.0342	18	1.7273	1.7007	0.0266
9	1.7158	1.7168	−0.0010	19	1.7224	1.7135	0.0089
10	1.7853	1.7938	−0.0085	20	1.6997	1.7118	−0.0121

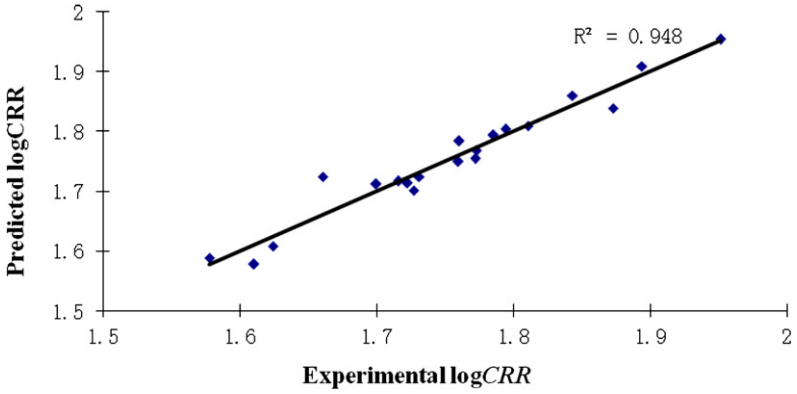


Figure 2. Experimental versus predicted according to the model in.

Table 4. Internal validation of the QSAR model

Training set	*N*	R^2 (fit)	S^2 (fit)	Test set	*N*	R^2 (pred)	S^2 (fit)
A + B	14	0.91	0.0007	C	6	0.98	0.0002
A + C	13	0.98	0.0003	B	7	0.86	0.0006
B + C	13	0.94	0.0010	A	7	0.98	0.0004
Average		0.94	0.0007			0.94	0.0004

three test sets still gives the satisfactory result, compared to the average R^2 (fit) of all the three training sets. Here, the obtained QSAR model obtained demonstrates the predictive power when threefold cross-validation is performed. The 'leave-one out' approach was performed in a similar manner to internal validation with the every fourth compound being put into an external test set, while the remaining compounds are included in the training set. QSAR models containing the same set of four descriptors were obtained with $R^2 = 0.95$. When the external set was tested, it gave the satisfactory $R^2_{\text{pred}} = 0.86$. These validation results indicate that the QSAR model obtained is statistically significant.

Descriptors involved in the qualitative model obtained describe the possible mechanisms by which terpenoid mosquito repellents work. The most statistically significant descriptor is boiling point. It has been previously shown that effective mosquito repellents should have an appropriate boiling point/vapor pressure to keep the duration of contact time of a mosquito with the repellent and to affect the olfactory chemosensilla of the mosquito.[25–27]

The second most statistically significant descriptor was dipole moment. Dipole moment is a good indicator of lipophilicity and hydrophobicity[28] with research suggesting an optimal dipole moment range for compounds having better repellency.[29] Notably, dipole moment only reflects the intrinsic polarity of a molecule and not the charge distribution of the molecular surface, an important factor in interactions between repellents and their receptor.

Oxygen-containing substituent groups play an important role in mosquito repellency.[30,31] Previous studies indicate that terpenoid mosquito repellents with two functional groups are biologically active.[32] All repellents in the present work have two functional/substituent groups: one negatively charged end containing either ester/ether bonds or an ethanol hydroxyl group and one positively charged end containing alkane groups. In order to explore their effects, charges on the end of func-

tional groups were summarized separately and counted as different descriptors that describe the polarity on each substituent group. The QSAR model reveals that the positive end is more favorable to receptor interactions. The magnitude of the positive charge characterizes the electrophilic nature of the group,[29] so the repellent–receptor interactions are most likely related to electrophilic interactions.

Structure fit is another primary concern in the binding of mosquito repellents to the active receptor center. Our results show that the molecular surface, instead of the molecular volume, is more statistically significant. This may imply that the binding of repellents by the receptor depends on the surface–surface contact instead of the lock-key fit of many biological processes.

The QSAR model was developed for the description of mosquito repellents with satisfactory statistical characteristics. The qualitative model obtained may reveal how structures of terpenoid mosquito repellents affect repellency. Such results are of great benefit to synthetic efforts to discover better compounds having practical uses.

Acknowledgments

Z.W., J.C., Z.S., and S.S. gratefully thank the Ministry of Science & Technology of China (No. 2006BAD06B10) for financial support. Z.W., J.C., Z.S., Z.J., and Z.H. acknowledge the National Natural Science Foundation of China (No. 30360085) and Z.W., J.C., Z.J., and Z.H. wish to thank the Fok Ying-Tong Education Foundation (No. 101031). J.S. thanks University of Michigan-Flint (RCAC award) for support.

References and notes

1. Hayes, W. J., Jr. *Pesticides Studied in Man*; Waverly Press, Inc.: Baltimore, MD, 1982.
2. Hallenbeck, W. H.; Cunningham-Burns, K. M. *Pesticides and Human Health*; Springer-Verlag: NY, 1985.
3. U.S. Environmental Protection Agency. Office of Pesticides and Toxic Substances. *N*,*N*-Diethyl-*m*-toluamide (DEET) Pesticide Registration Standard, 1980.
4. New York State Department of Health, Bureau of Toxic Substances Assessment/New York State Department of Environmental Conservation. Hearing Notice for DEET, 1991.
5. The Agrochemicals Handbook. The Royal Society of Chemistry. The University, Nottingham, England, 1983.
6. Partridge, M. S.; Gardner, R. D.; Rutz, D. A. Non-agricultural Pesticide Use in the Wallkill/Rondout Watershed Region of New York State for 1991, 1994.
7. Wang, Z. *Dissertation* ; Chinese Academy of Forestry: Beijing, 2005.
8. Wang, Z.; Jiang, Z.; Song, Z. *Chin. J. Hyg. Insect. Equip.* **2004**, *10*, 37.
9. Thorsell, W.; Tunon, H. *Phytomedicine* **1998**, *5*, 307.
10. Bhattacharjee, A. K.; Gupta, R. K. *J. Am. Mosq. Control Assoc.* **2005**, *21*, 23.
11. Hwang, Y. S.; Schultz, G. W.; Mulla, M. S. *J. Chem. Ecol.* **1984**, *10*, 145.
12. Suryanarayana, M. V. S.; Pandey, K. S.; Prakash, ShriS.; Raghuveeran, C. D.; Dangi, R. S.; Swamy, R. V.; Rao, K. M. *J. Pharm. Sci.* **1991**, *80*, 1055.
13. Katritzky, A. R.; Dobchev, D. A.; Tulp, I.; Karelson, M.; Carlson, D. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2306.
14. CODESSA 2.7.10. Katritsky, A.; Karelson, M.; Lobanov, V. S.; Dennington, R.; Keith, T. Semichem, Inc., Shawnee, KS, 2004.
15. Wang, Z.; Chen, J.; Song, Z.; Jiang, Z.; Han, Z.; Chen, C. *Chem. Industry Forest Prod.* **2007**, *27*, 1.
16. Chen, J.; Xiao, Z.; Wang, Z. *Jiangxi Nongye Daxue Xuebao* **2006**, *28*, 953.
17. Wang, Z.; Xiao, Z.; Chen, J. *Huaxue Shijie* **2004**, *45*, 89.
18. Wang, Z.; Chen, J.; Song, Z.; Xiao, Z. *J. Fujian Agric. Forestry Univ. (Nat. Sci. Edition)* **2007**, *36*, 385.
19. Skinner, W. A.; Johnson, H. L. *Med. Chem.* **1980**, *11*, 277.
20. Davis, E. E. *J. Med. Entomol.* **1985**, *22*, 237.
21. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C., Pople, J. A. Gaussian 03, Revision D.01. Gaussian, Inc.: Wallingford, CT, 2004.
22. AMPAC 8. Dewar, M. J. S.; Holder, A. J.; Dennington, R. D., II ; Liotard, D. A.; Truhlar, D. G.; Keith, T. A.; Millam, J. M.; Harris, C. D. Semichem, Inc.: Shawnee, KS, 2004.
23. SciFinder Scholar, American Chemical Society, 2008.
24. Zhang, G.; Fang, B. *Process Biochem.* **2006**, *41*, 552.
25. Davis, E. E.; Rebert, C. S. *J. Econ. Entomol.* **1976**, *105*, 1058.
26. Rayner, H. B.; Wright, R. H. *Can. J. Entomol.* **1966**, *98*, 76.
27. Johnson, H. L.; Skinner, W. A.; Skidmore, D.; Maibach, H. I. *J. Med. Chem.* **1968**, *11*, 1265.
28. McIver, S. B. *J. Med. Chem.* **1981**, *11*, 357.
29. Ma, D.; Bhattacharjee, A. K.; Gupta, R. K.; Karle, J. M. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 1.
30. Alexander, B. H.; Beroza, M. *J. Econ. Entomol.* **1963**, *56*, 58.
31. Roadhouse, L. A. O. *Can. J. Zool.* **1953**, *31*, 535.
32. Jiang, Z.; Zheng, Z.; Zhao, X. *Application and Research of Hygienic Insecticides & Equipments*; Nangjing University Press: China, 2001.