J Mol Model (2011) 17:681–693
DOI 10.1007/s00894-010-0765-x

ORIGINAL PAPER

Isolation, biological evaluation and 3D-QSAR studies of insecticidal/narcotic sesquiterpene polyol esters

Shao-peng Wei · Zhi-qin Ji · Hui-xiao Zhang · Ji-wen Zhang · Yong-hua Wang · Wen-jun Wu

Received: 20 April 2010 / Accepted: 20 May 2010 / Published online: 8 June 2010

Abstract For the first time, a set of (43) natural sesquiterpene polyol esters isolated from the root bark of *Celastrus angulatus* Maxim and *Euonymus japonicus* Thunb were subjected to 3D-QSAR comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) studies, with the aim of proposing novel sesquiterpene-based compounds with optimal narcotic or insecticidal activities. The established 3D-QSAR models exhibit reasonable statistical quality and prediction capabilities, with internal cross-validated Q^2 values of ~0.5 and external predicted R^2 values of >0.9, respectively. The relative contributions of the steric/electrostatic fields of the 3D-QSAR models show that the electronic effect governs the narcotic activities of the molecules, but the hybrid effect of the electrostatic and hydrophobic interactions is more influential in the insecticidal activities of the compounds. These findings may have valuable implications for the development of novel natural insecticides.

Keywords Sesquiterpene polyol esters · Isolation · Insecticide · 3D-QSAR

Shao-peng Wei, Zhi-qin Ji and Hui-xiao Zhang have contributed equally.

S.-p. Wei · Z.-q. Ji · H.-x. Zhang · J.-w. Zhang · Y.-h. Wang (✉) · W.-j. Wu (✉)
Institute of Pesticide Science,
Northwest Agricultural & Forestry University,
Shaanxi 712100, China
e-mail: yh_wang@nwsuaf.edu.cn
e-mail: wuwenjun@nwsuaf.edu.cn

Y.-h. Wang
Laboratory of Pharmaceutical Resource Discovery,
Dalian Institute of Chemical Physics,
the Chinese Academy of Sciences,
#457 Zhongshan Road,
Dalian 116023, China

Y.-h. Wang
School of Chemical Engineering,
Dalian University of Technology,
#158 Zhongshan Road,
Dalian 116312, China

Introduction

As they are a valuable source of new lead compounds and chemical entities in the agrochemical and pharmaceutical industries, natural products account for 30% of international drug sales [1]. Natural products play such a pivotal role in drug discovery that they offer a great diversity of chemical scaffolds for finding new lead compounds and optimizing their efficacies to a certain extent [1]. Compared with recombinant proteins and peptides, one kind of natural product—plant-derived drug candidates—has the advantage of possessing low molecular mass, thus leading to more favorable compliance and bioavailability. Many valuable drugs have been isolated from plant sources, including aspirin, morphine, reserpine, digitalis, and anticancer agents such as vincristine, vinblastine, and taxol [2]. In addition, various kinds of natural products have been used as insecticidal agents that target a wide variety of insect pests, such as the spinosyns [3], pyrethroids, azadirachtin, alkaloids, fluvalinate, etc. [4–6]. Among these compound classes, sesquiterpenes from Celastraceae are promising insecticides, some of which have already been adopted to protect crops from insects in China [7–9].

One member of the Celastraceae family, the Chinese bittersweet, *Celastrus angulatus*, is a traditional insecticidal plant that is widely distributed across the Yellow River and Yangtze River basins of China. In order to find new bioactive natural products that can be used as insecticides, our group has carried out research on this particular plant for more than 20 years, since 1985 [10, 11]. In our studies, the organic

solvent extract of the root bark of *C. angulatus* exhibited antifeedant, narcotic, and insecticidal activities against several insect species [11]. Our subsequent analysis of the chemical compositions of this plant further showed that the active ingredients are a series of β-dihydroagarofuran sesquiterpenoids, whose core structure is a 5,11-epoxy-5α,10 β-eudesman-4-(12)-ene skeleton that comprises A and B rings in the form of an axially dimethylated *trans*-decalin bicycle with a 1,3-diaxially fused Me_2C–O bridge that constitutes the tetrahydrofuranyl C-ring (Fig. 1) [12].

Sesquiterpene polyol esters with a β-dihydro-agarofuran skeleton are typical secondary metabolites in Celastraceae, and these have attracted considerable attention due to their interesting structures [12–14] and their wide-ranging biological properties, including immune-suppressive [15, 16], cytotoxic [17], anti-HIV [18], reversing multi-drug resistance (MDR) phenotype [13, 19–21], antitumor [22, 23], and insect narcotic and insecticidal [9, 24–29] activities.

Previously, our laboratory isolated some sesquiterpene compounds and alkaloids from the root bark of *C. angulatus* that demonstrated excellent bioactivity against many species of insects from the family Lepidoptera, such as the *Pieris rapae, Plutella xylostella* and *Mythimna separata* [10]. In the present work, 43 analogs of sesquiterpene polyol esters were isolated from two plant species (*C. angulatus* and *E. japonicus*), and two distinct biological activities (narcotic and insecticidal) against fourth-instar larvae of *M. separata* were investigated. Additionally, in order to explore the structural elements that are crucial to the narcotic and insecticidal activities of these compounds, a three-dimensional quantitative structure–activity relationship (3D-QSAR) study using an extension of the comparative molecular field analysis (CoMFA) methodology and a comparative molecular similarity indices analysis (CoMSIA) was performed.

Experimental section

Isolation

All natural sesquiterpenes were obtained from two Celastraceae species. Compounds **1–39** were isolated from the root bark of *C. angulatus* Max.; compounds **40–43** were isolated from the root bark of *E. japonicus* (Fig. 2). The general isolation procedure is shown in Fig. 3.

The dried and pulverized root bark (5.5 kg) of *C. angulatus* was extracted three times under reflux (4 h each time) with analytical grade benzene. The dried extracted components were then re-extracted twice with methanol, after which the methanolic extracts were combined and concentrated to yield a yellow semisolid residue (500 g). This crude extract was adsorbed onto a D101 macroporous resin column (15.0×150 cm) and gradient eluted with methanol–water (1:9~9:1); 200 fractions of ca. 500 ml each were collected. After analysis with thin-layer chromatography, similar fractions were combined to afford ten fractions. Then the low-polar fractions were successively chromatographed on a silica gel (200–300 mesh) column, eluting with mixtures of petroleum ether–acetone–methanol of increasing polarity. Using methanol–water as the eluant, the polar parts were isolated on an RP flash column. In this way, i.e., by several chromatographies on silica gel (petroleum ether–acetone or acetone–methanol), RP flash (methanol–water) and pre-HPLC, compounds **2**, **4**, **6–8**, **10**, **12–13**, **15**, **17–19**, **21**, **23**, **28–33**, **35–39** were isolated (Fig. 3A). All of the chemicals were subsequently identified by UV, IR, ESI-MS, ^{1}H and ^{13}C-NMR [24–27, 30, 31].

The dried and pulverized root bark (2.0 kg) of *E. japonicus* was extracted three times under reflux (4 h each time) with methanol. The extracts obtained were combined and concentrated to yield a yellow semisolid residue (150 g). This crude extract was successively chromatographed on a silica gel (200–300 mesh) column using mixtures of petroleum ether–acetone–methanol of increasing polarity as eluent. After several chromatographies on silica gel (petroleum ether–acetone or acetone–methanol) and pre-HPLC, compounds **40–43** were isolated (Fig. 3B). Their structures were identified based on UV, IR, ESI-MS, ^{1}H and ^{13}C-NMR spectroscopic evidence.

The fresh root bark of *C. angulatus* was collected on Qinling Mountain, Taibai County, Shaanxi Province, People's Republic of China, in October 2006. It had a reddish color and was authenticated by Prof. Hua Yi of the College of Life Sciences, Northwest Agricultural & Forestry University. The bark was air-dried and then cut into small pieces. The voucher specimens (samples no. NWAU2006-A18) were deposited at the College of Life Sciences, Northwest Agricultural & Forestry University.

The fresh root bark of *E. japonicus* was collected from a nursery in Yangling, Shaanxi Province, People's Republic

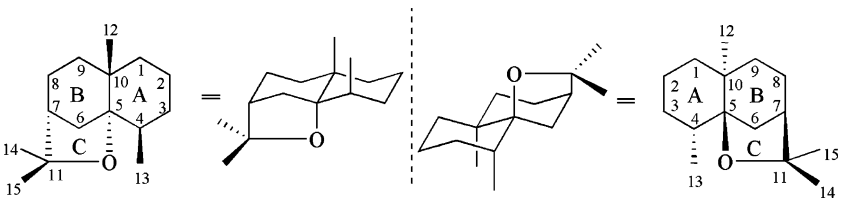

Fig. 1 General skeleton of β-dihydroagarofuran

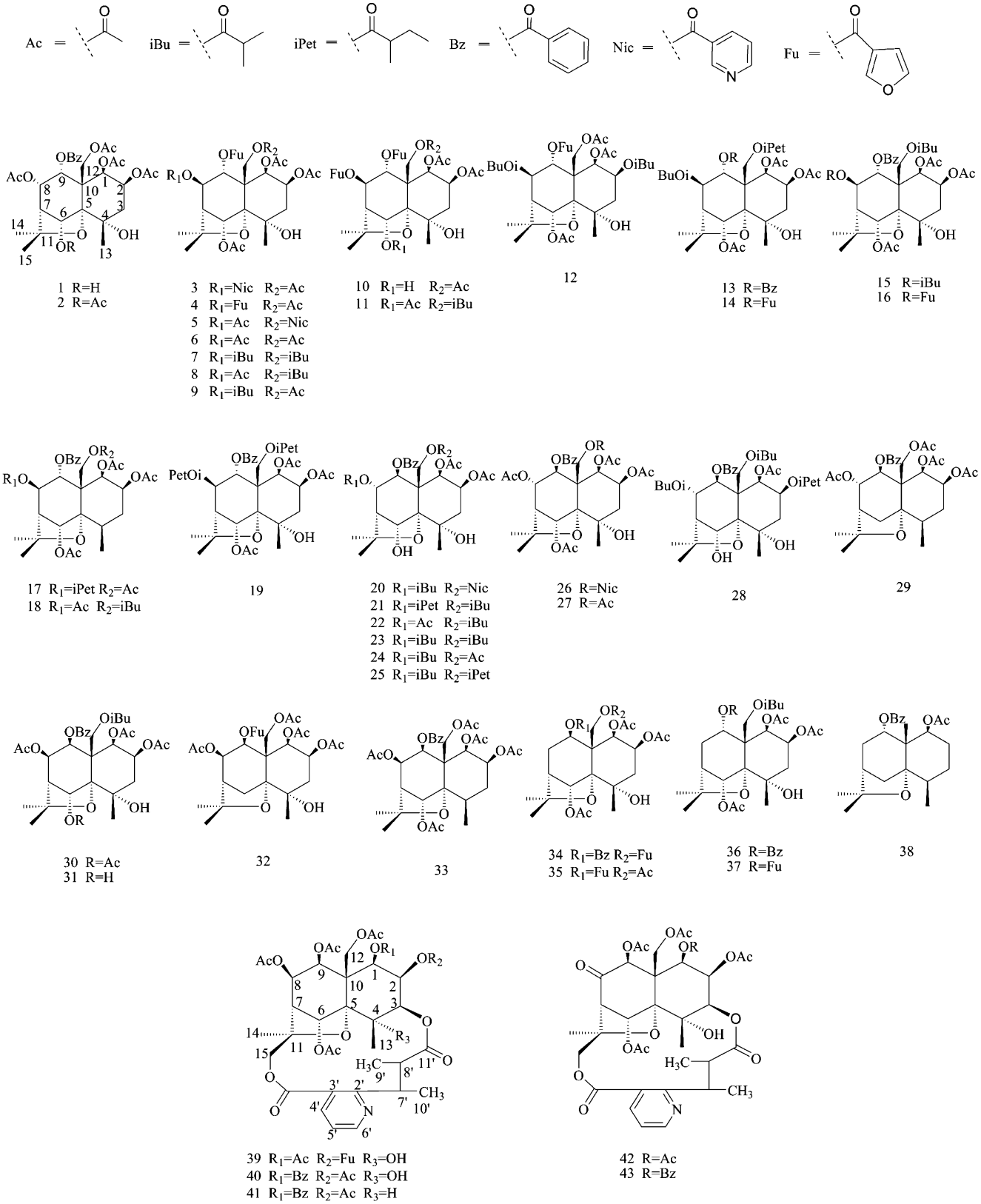

Fig. 2 Structures of the sesquiterpenes

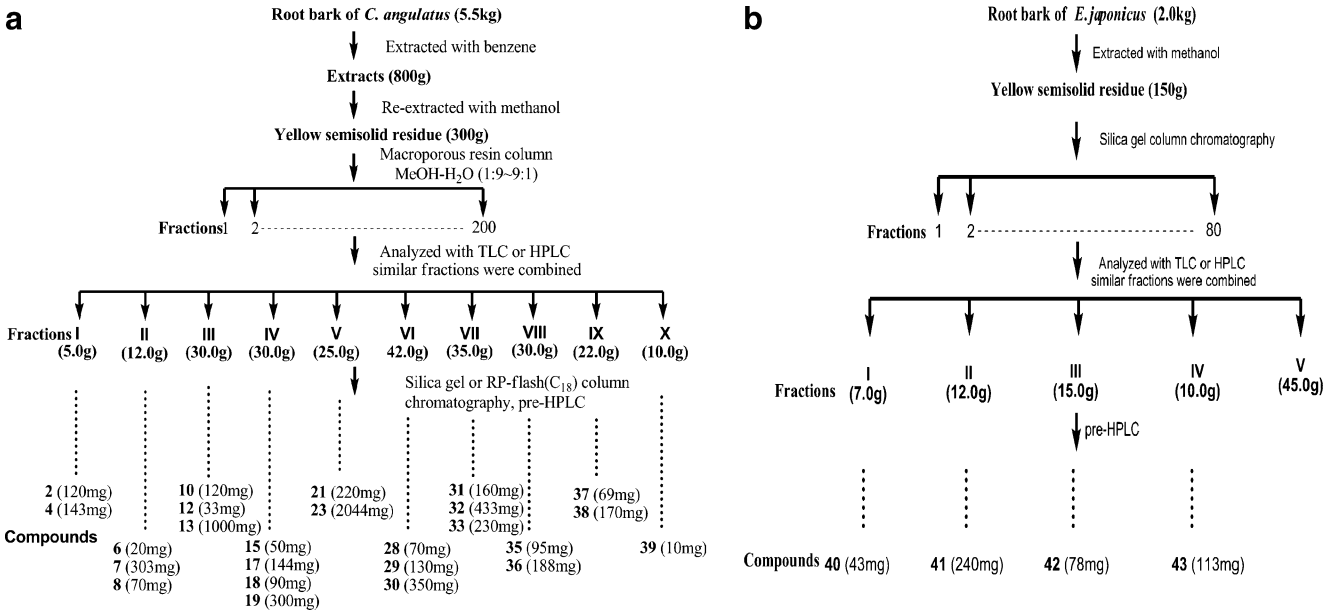

Fig. 3 Isolation of *C. angulatus* (**A**) and *E. japonicus* (**B**)

of China, in October 2006. It had a yellowish color and was authenticated by Prof. Hua Yi of the College of Life Sciences, Northwest Agricultural & Forestry University. The bark was air-dried and then cut into small pieces. The voucher specimens (samples no. NWAU2006-A19) were deposited at the College of Life Sciences, Northwest Agricultural & Forestry University.

During the purification process, unless otherwise stated, all solvents were of analytical reagent (AR) grade. All fractions were monitored by thin-layer chromatography (TLC) with E. Merck $60F_{254}$ silica gel plates. Flash column chromatography was performed with the indicated solvents on C_{18} chromatography silica gel (particle size 15 μm, Fuji Silysia Chemical Ltd.). Melting points were measured on a Yanagimoto apparatus, and were uncorrected. Proton NMR spectra were recorded on a Bruker Avance 500 MHz spectrometer using $CDCl_3$ as solvent and TMS as internal standard. Compounds were purified with a Shimadzu 6AD HPLC apparatus equipped with a C_{18} preparative column (20×250 mm, 10 μm), MeOH-H_2O (5:5–8:2) as eluent, and UV detection at 230 nm. The ionization method was positive ion electrospray with an ESI model, and the molecular scan range was 100–1000 amu, performed on a Thermo Finnigan LCQ Advantage MAX LC/MS mass spectrometer. Samples were supplied as 0.5 mg ml^{-1} in methanol, with 10 μL injected on a partial loop fill.

Bioassay

Mythimna separata was continuously maintained in our laboratory at a constant temperature of 27 °C under a 16L:8D photoperiod. The larvae were reared on wheat leaves.

The bioactivities of all compounds against the fourth-instar larvae of *M. separata* were tested with wheat leaves (0.5×0.5 cm) treated with 1 μL of an acetone solution of the test compounds at concentrations of 5, 10, 15, 20 and 25 mg·ml^{-1}, respectively. For each bioassay, three treated leaves were placed alternately in a Petri dish (Ø 6 cm) with five larvae (weight 14–20 mg) and kept in a growth chamber in the dark at 27 °C and 60% relative humidity. The larvae were fed with the treated leaves for 24 h, and each sample was repeated ten times with celangulin V as the positive control and clean leaves as the negative control. Then the eaten area of a leaf was measured under a binocular microscope and the dose of the chemical was calculated. After 24 h, the numbers of the narcotized (Fig. 4A) (symptoms: the larvae were immobilized, soft, could not move and had completely lost their ability to respond [27]) and dead (Fig. 4B) (symptoms: the larvae were knocked down, serious loss of body fluid and finally death [27]) larvae were recorded, and the toxicity was ascertained by estimating the median narcotic dose (ND_{50}, producing 50% narcosis) or the median lethal dose (LD_{50}, killing 50%) of the test sample [27].

Molecular modeling

In this work, the LD_{50} and ND_{50} (μM g^{-1}) were converted to pLD_{50} ($-\log LD_{50}$) and pND_{50} ($-\log ND_{50}$) values, which were used as dependent variables in the CoMFA and CoMSIA QSAR analyses. Selection of the training and the

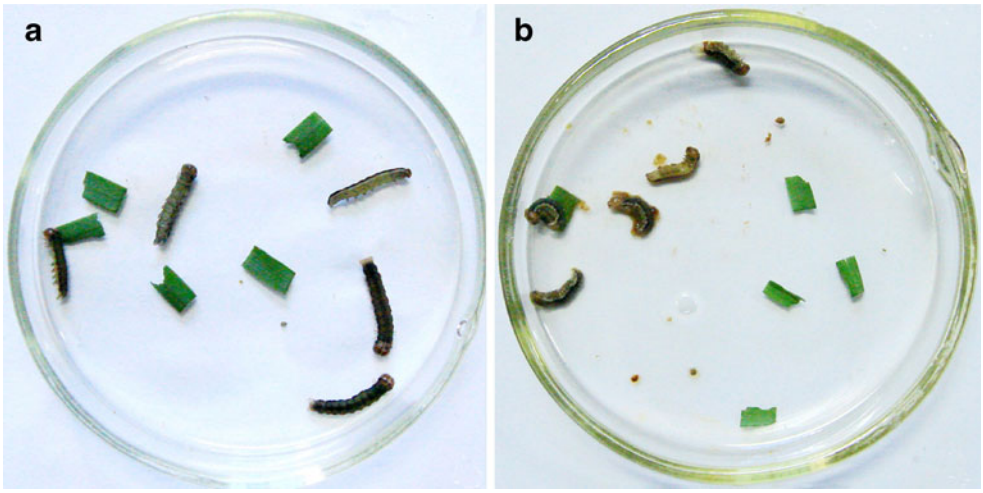

Fig. 4 Narcotized (**A**) and dead (**B**) fourth-instar larvae of *Mythimna separata*

test sets was done by considering the fact that the test compounds represent structural diversity and a range of biological activities similar to that of the training set (Tables 1 and 2). The mean activities of the training and test set compounds in the narcotic model (pND_{50}) were 0.33 and 0.24, respectively, and those in the insecticidal model (pLD_{50}) were 0.17 and 0.16, respectively, which confirmed that the test set was truly representative of the training set. The structures of the compounds in the training and test sets are shown in Fig. 2.

All molecular modeling and calculations were carried out using Sybyl 6.9 (Tripos Inc.) on the Redhat Linux platform. Energy minimizations for narcotic and insecticidal compounds were performed using the Tripos force field [32, 33] and Gasteiger–Hückel charges with other conjugate gradient minimization algorithm with a convergence criterion of 0.05 kcal mol^{-1} $Å^{-1}$ and a total of 200 maximization iterations. Based on the atom-by-atom superimposition of selected atoms, all molecules were aligned. In narcotic and insecticidal modeling, the most active compounds (**43** and **16**) were used as the template molecules, respectively, and all other compounds were aligned on the basis of the maximum common structure shown in Fig. 5. The respective alignment results for the narcotic and insecticidal models are shown in Fig. 6.

CoMFA and CoMSIA 3D-QSAR models

In order to derive the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with a grid spacing of 2 Å in the *x*, *y* and *z* directions was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp^3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard–Jones 6–12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. Values of the steric and electrostatic fields were truncated at a default value of 30 kcal mol^{-1}. This signifies that any steric or electrostatic field value that exceeds this value will be replaced with 30 kcal mol^{-1}, thus making the fields close to the center of any atom plateau. The CoMFA steric and electrostatic fields were scaled by the CoMFA-STD method [34] in Sybyl. In addition, MLog*P*, as calculated by DRAGON (http://www.talete.mi.it/help/dragon_help/), was also added to observe its effect on the correlation with compound activity.

CoMSIA descriptors were calculated according to Klebe [35] with the same lattice box as that used for the CoMFA calculations, with a regularly placed grid of 2.0 Å, and a C^{1+} probe atom with a radius of 1.0 Å as implemented in

Table 1 Actual vs. predicted pND_{50} values of the optimal CoMSIA model in the narcotic study

Compound	Actual pND_{50}	Predicted pND_{50}	Residual
Training set			
2	0.35	0.22	0.13
4	0.13	−0.08	0.21
7	−0.01	−0.08	0.07
15	0.04	0.21	−0.17
35	−0.34	−0.11	−0.23
42	0.8	1.04	−0.24
43	1.4	1.18	0.22
Test set			
1	0.17	0.27	−0.1
11[a]	0.74	0.03	0.71
18	0.08	0.1	−0.02
41	0.47	0.68	−0.21

[a] Outlier compound

Table 2 Actual vs. predicted pLD_{50} values and residuals of the optimal CoMSIA model in the insecticidal study

Compound	Actual pLD_{50}	Predicted pLD_{50}	Residual
Training set			
3	−0.23	−0.19	−0.04
5	−0.15	−0.18	0.03
8	−0.02	−0.03	0.01
9	0.12	0.07	0.05
10	−0.15	−0.1	−0.05
12	0.09	0.03	0.06
13	0.12	0.14	−0.02
14	0.32	0.39	−0.07
16	1.15	1.15	−0.002
17	−0.22	−0.19	−0.03
19	0.12	0.11	0.01
21	0.34	0.36	−0.02
24	0.19	0.19	−0.001
25	0.68	0.65	0.03
26	0.08	0.06	0.02
27	0.55	0.58	−0.03
28	−0.01	0.05	−0.06
29	0.02	−0.01	0.03
30	−0.03	−0.01	−0.02
31	0.02	−0.01	0.03
32	−0.38	−0.39	0.01
33	0.27	0.25	0.02
34	0.35	0.34	0.01
37	0.37	0.32	0.05
38	−0.25	−0.24	−0.01
39	0.37	0.36	0.01
40	0.92	0.92	−0.004
Test set			
6	−0.25	−0.18	−0.07
20	0.55	0.45	0.1
22	0.07	0.05	0.02
23	0.33	0.23	0.1
36	0.14	0.17	−0.03

Sybyl. CoMSIA similarity indices (A_F) for molecule j with atoms i at the grid point q were determined using Eq. 1:

$$A_{F,K}{}^{q}(j) = -\sum_{i} \omega_{probe,k}\omega_{ik}e^{-\alpha r_{iq}^{2}}, \quad (1)$$

where i is the summation index over all atoms of the molecule j under investigation; ω_{ik} is the actual value of the physicochemical property k of atom i; $\omega_{probe,k}$ is the probe atom with radius 1.0 Å, charge +1, hydrophobicity +1, hydrogen bond donating +1, hydrogen bond accepting +1; and r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecule. k represents the steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor properties. A Gaussian-type distance dependence was used between the grid point q and each atom i of the molecule. A default value of 0.3 was used as the attenuation factor α. In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors were derived from atomic partial charges, and the hydrophobic fields were derived from atom-based parameters developed by Viswanadhan et al. [36]. The H-bond donor and acceptor indices were obtained by a rule-based method based on experimental results [37].

In the partial least-square (PLS) regression analysis, CoMFA and CoMSIA descriptors were used as independent variables, and the pLD_{50} or pND_{50} values were used as dependent variables [38] to derive 3D-QSAR models. The predictive value of these models was evaluated first by leave-one-out (LOO) cross-validation, the R^2_{cv} coefficient of which was calculated by Eq. 2:

$$R^2{}_{cv} = 1 - \frac{\sum \left(Y_{predicted} - Y_{actual}\right)^2}{\sum \left(Y_{actual} - Y_{mean}\right)^2}, \quad (2)$$

where $Y_{predicted}$, Y_{actual}, and Y_{mean} are the predicted, actual, and mean values of the target property (pLD_{50} or pND_{50}), respectively. $\sum \left(Y_{predicted} - Y_{actual}\right)^2$ is the predictive residual sum of squares (PRESS).

Based on the training set compounds, the optimal number of components (ONC) derived from the cross-validated PLS analysis were used to generate the final QSAR model. To test the validity of the model, an external set of compounds with known activities that were not used in model generation (the test set) were predicted. The predictive R^2 (denoted by R^2_{pred}) was calculated by Eq. 3:

$$R^2{}_{pred} = 1 - (\mathrm{PRESS}/\mathrm{SD}), \quad (3)$$

where SD is the sum of the squared deviations between the biological activities of the test set molecules and the mean activity value of the training set molecules. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

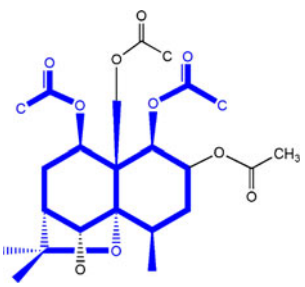

Fig. 5 All of the atoms of this scheme were used in the alignment of the narcotic model, and the atoms shown in *blue* were used for the insecticidal model

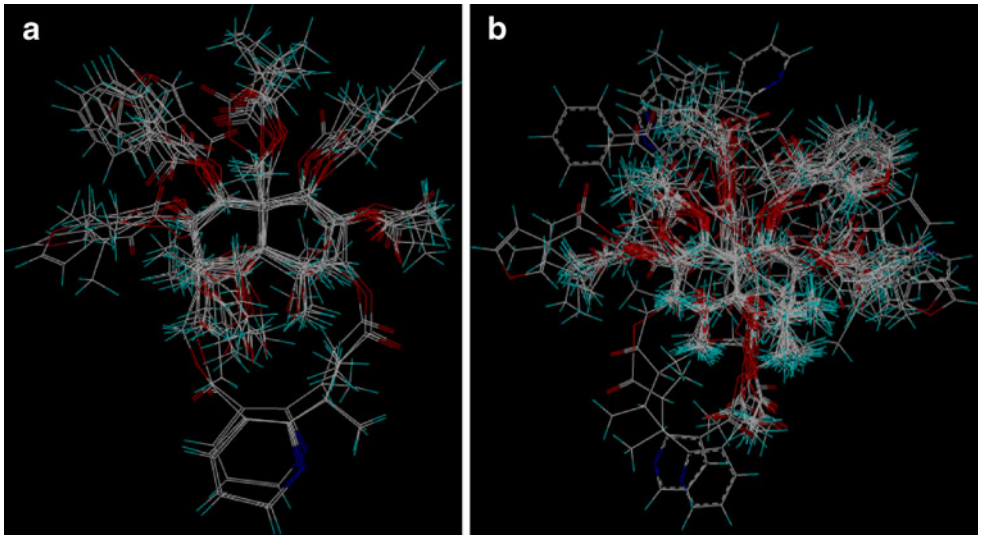

Fig. 6 Structural alignments of the 11 compounds in the narcotic model (**A**) and the 32 compounds in the insecticidal model (**B**)

Results and discussion

Biological activity

Celastrus angulatus (Celastraceae) is an insecticidal plant that is widely distributed across China. Its insecticidal active ingredients are sesquiterpenoid polyol esters with a dihydroagarofuran structure [39]. The compounds isolated from this plant have demonstrated light, heat and pH stability, and exert unique biological activities involving narcotic and toxic poisoning actions [39]. A 0.2% celangulins emulsifiable concentrate has been developed from this plant and used for insect control in vegetables, tea and city virescence [39]. However, the effective components of the plant—the specific compounds that determine its insecticidal activity—have not yet been fully identified, and so this is the aim of these studies.

Previously, celangulin IV (i.e., compound **11** in this work, ND_{50}=0.18 μM g^{-1}) and celangulin V (compound **23**, LD_{50}=0. 47 μM g^{-1}) demonstrated significant narcotic and insecticidal activities against the fourth-instar larvae of *Mythimna separata*, respectively [39]. In our study, compounds **43** (ND_{50}=0.04 μM g^{-1}) and **16** (LD_{50}= 0.07 μM g^{-1}) were found to be about five times more effective than the previous two potential drugs, indicating the potential of these newly isolated compounds to be more effective drug candidates.

In order to deeply explore the structural determinants that affect the activities of these compounds, and also to develop more potent compounds for synthesis, 3D-QSAR studies using CoMFA and CoMSIA methods were carried out on these natural β-dihydroagarofuran sesquiterpene polyesters.

Narcotic model

The steric and electrostatic fields were applied as molecular descriptors in the CoMFA analysis, while the steric, electrostatic, and hydrophobic fields were used in the CoMSIA analysis, because our calculations (data not shown) indicated that the H-bond donor and acceptor descriptors were not beneficial to the biological activities. Four different models were generated by applying Mlog*P* as an additional descriptor in CoMFA and CoMSIA analysis in order to study its effects on the correlation with the molecular activity. The results obtained from the four models are summarized in Table 3.

For the 11 compounds, a cross-validated coefficient (R^2_{cv}) of 0.36 and a non-cross-validated coefficient (R^2_{ncv}) of 0.92 with two optimum components for the CoMFA model, an R^2_{cv} of 0.26 and an R^2_{ncv} of 0.97 with three optimum components for the CoMFA-Mlog*P* model, an R^2_{cv} of 0.29 and an R^2_{ncv} of 0.98 with three optimum components for the CoMSIA model, and an R^2_{cv} of 0.02 and an R^2_{ncv} of 0.97 with three optimum components for the CoMSIA-Mlog*P* model were observed. Generally, in 3D-QSAR CoMFA and CoMSIA studies, an R^2_{cv} of 0.4 is considered to be statistically significant [40]. Thus, these modest R^2_{cv} values indicated that some compounds may be outliers. After dropping compound **11** from the training set, 3D-QSAR studies were reperformed on the remaining ten compounds. As a result, an increased R^2_{cv} value of 0.43 with an ONC of 2 for the CoMFA model, an R^2_{cv} of 0.18 with an ONC of 3 for the CoMFA-Mlog*P* model, an R^2_{cv} of 0.47 with an ONC of 1 for the CoMSIA model, and an R^2_{cv} of 0.15 with an ONC of 3 for the CoMSIA-Mlog*P* model were attained (Table 3). These results were

Table 3 Summary of 3D-QSAR results for the narcotic model

10-compound model					11-compound model				
PLS statistics	CoMFA	CoMFA +Mlog*P*	CoMSIA	CoMSIA +Mlog*P*	PLS statistics	CoMFA	CoMFA +Mlog*P*	CoMSIA	CoMSIA +Mlog*P*
R^2_{ncv}	0.94	0.97	0.87	0.97	R^2_{ncv}	0.92	0.97	0.98	0.97
SEE	0.18	0.14	0.23	0.15	SEE	0.18	0.13	0.09	0.14
F values	30.43	34.89	34.98	28.17	*F* values	29.82	45.11	87.34	38.23
R^2_{cv}	0.43	0.18	0.47	0.15	R^2_{cv}	0.36	0.26	0.29	0.02
SEP	0.18	0.22	0.13	0.20	SEP	0.14	0.20	0.17	0.15
R^2_{pred}	0.99	0.96	0.99	0.99	R^2_{pred}	0.98	0.87	0.87	0.97
ONC	2	3	1	3	ONC	2	3	3	3
Contribution					Contribution				
Steric	0.563	0.522	0.158	0.146	Steric	0.563	0.511	0.130	0.136
Electrostatic	0.437	0.421	0.522	0.486	Electrostatic	0.437	0.400	0.531	0.464
Hydrophobic			0.319	0.326	Hydrophobic			0.340	0.329
Mlog*P*		0.057		0.042	Mlog*P*		0.088		0.072

R^2_{ncv}, non-cross-validated correlation coefficient; SEE, standard error of estimate. R^2_{cv}, cross-validated correlation coefficient using the leave-one-out methods; SEP, standard error of prediction. R^2_{pred}, predicted correlation coefficient for the test set of compounds; ONC, optimal number of components

considerably more accurate than those obtained using the whole dataset.

There are several factors that may account for the outlier status of compound **11**, including unique structural differences, different binding conformations, and a higher residual between the observed and predicted biological activity of an inhibitor. In the present study, compound **11** was classified as an outlier since the prediction error (0.71) for this compound in the sample is over three times the overall variance in the data (0.21).

Although compound **11** is very similar in structure to **4**, with the only difference occurring at the C12 position (where compound **11** has an isobutyryloxy and compound **4** an acetoxy), omitting compound **11** from the training set results in an obvious improvement of the model in our calculations. This may be due to several reasons, such as different binding conformations of the compounds, or the compound dataset being too small, etc., but this requires further experimental validation.

It was found that adding Mlog*P* to the CoMFA and CoMSIA analyses caused a reduction in R^2_{cv} and contributed 5.7% to the CoMFA-Mlog*P* model, as well as 4.2% to the CoMSIA-Mlog*P* model. The CoMSIA model was selected by comparison to be the optimal one, based on its better statistics, particularly the fact that it had the highest R^2_{cv} (0.47) and R^2_{pred} (0.99) values. Contour maps of the optimal model are shown in Fig. 7A–C.

Validation of the 3D-QSAR models

In 3D-QSAR modeling, validation is always a crucial step. It has been found that the widely accepted LOO cross-validated R^2_{cv} cannot fully assess the predictive abilities of the QSAR models [41], since some models based on the training set with randomized affinities were found to have high R^2_{cv} values but low predictive powers in an external test set. This could be interpreted as being due to a chance correlation or structural redundancy [42]. Thus, in order to validate the QSAR model, the predictive power of the optimal model was investigated using a test set (about 30% of all compounds) that was not included in model generation. The optimal model shows excellent prediction (R^2_{pred}=0.99) for the tested compounds, which indicates that the derived model is sufficient in terms of statistical significance and actual predictive ability. Graphs of the actual versus the predicted pND_{50} values of both the training and test sets using the optimal model are shown in Fig. 8.

CoMSIA contour maps

In the steric fields of CoMSIA, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80%- and 20%-level contributions, respectively. Similarly, in the electrostatic fields of CoMSIA, the blue

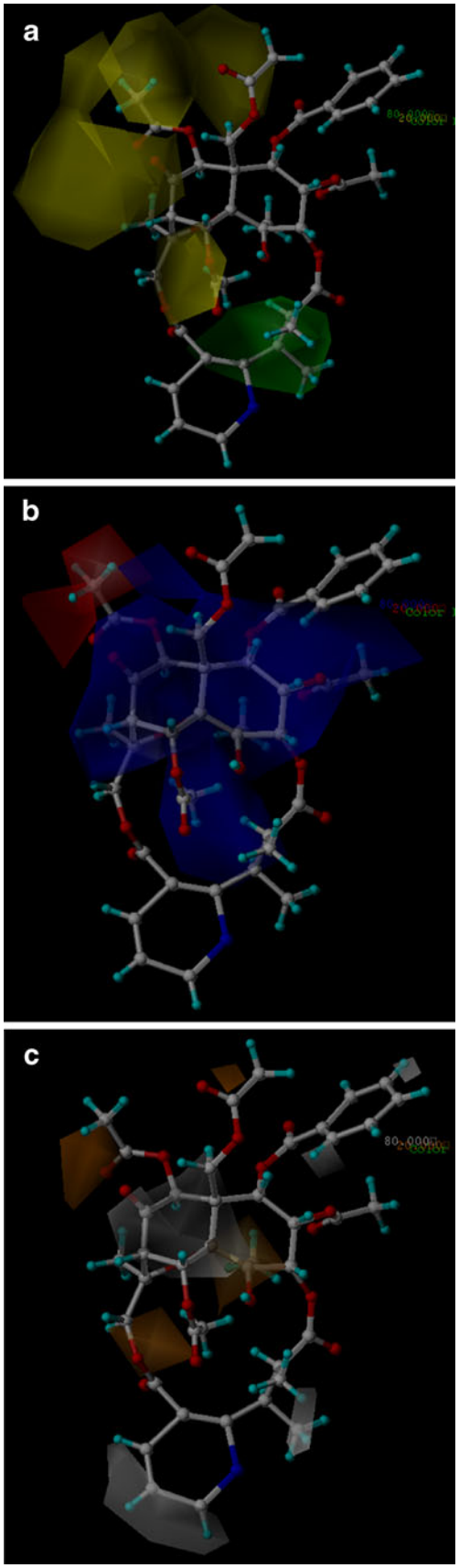

Fig. 7 CoMSIA stdev*coeff steric (**A**), electrostatic (**B**) and hydrophobic (**C**) contour maps for the narcotic model. The color code is as follows: in **A**, *green and yellow contours* indicate favorable and unfavorable bulky groups, respectively; in **B**, *blue and red contours* indicate favorable and unfavorable electropositive groups, respectively; in **C**, *white and orange contours* indicate favorable and unfavorable hydrophobic groups, respectively. The most potent compound, **43**, is displayed as a reference

(electropositive charge favorable) and red (electronegative charge favorable) contours represent 80%- and 20%-level contributions, respectively. In the hydrophobic field of CoMSIA, the white (hydrophobic favorable) and orange (hydrophobic unfavorable) contours represent 80%- and 20%-level contributions, respectively.

The CoMSIA contour map of steric contribution is shown in Fig. 7A. The most active compound, **43**, is displayed to aid the visualization of the map. A large yellow contour map near the C8 position of the B ring indicates that substituents at this region have unfavorable steric interactions. This is confirmed by the fact that for compound **43**, a small substituent (=O group) is located at this position, while for other less active compounds such as compound **15**, bulky substituents (isobutyryloxy group) occur in this region. The green contour near the C2′, 3′, 7′, and 8′ positions suggests the need for a bulky substituent in this area to enhance the biological activity, an observation that is consistent with experimental findings, such as those for the potent compounds **41**–**43**. The large yellow contour near the C9 position of the B ring indicates that a bulky substituent at this position would decrease biological activity, which is in accord with the fact that compounds **7** and **18**, which both have a bulky substituent at this position, possess pND_{50} values of <0.1.

Figure 7B shows the electrostatic contour map of the optimal model, with compound **43** used as the reference molecule. A red-colored region is clearly observed close to

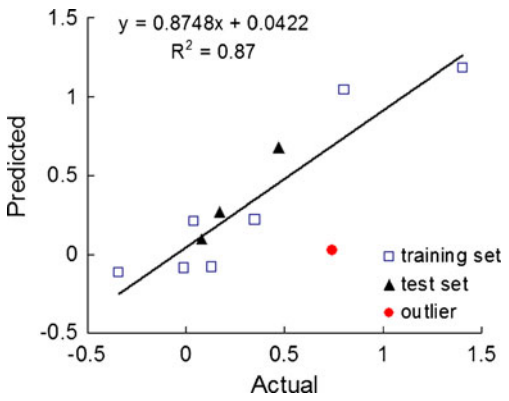

Fig. 8 Plot of actual versus predicted pND_{50} values for the training and test compounds based on the optimal CoMSIA model for the narcotic study

Table 4 Summary of the 3D-QSAR results for the insecticidal model

PLS statistics	CoMFA	CoMFA+MlogP	CoMSIA	CoMSIA+MlogP
R^2_{ncv}	0.99	0.99	0.99	0.98
SEE	0.04	0.03	0.04	0.05
F values	501.42	463.48	369.60	202.57
R^2_{cv}	0.39	0.24	0.46	0.31
SEP	0.15	0.20	0.07	0.13
R^2_{pred}	0.87	0.85	0.98	0.97
ONC	5	6	6	6
Contribution:				
Steric	0.770	0.742	0.210	0.198
Electrostatic	0.230	0.218	0.388	0.349
Hydrophobic			0.402	0.396
MlogP		0.040		0.057

R^2_{ncv}, non-cross-validated correlation coefficient; SEE, standard error of estimate. R^2_{cv}, cross-validated correlation coefficient using the leave-one-out methods; SEP, standard error of prediction. R^2_{pred}, predicted correlation coefficient for the test set of compounds; ONC, optimal number of components

the substituent at the C9 position, suggesting a high demand for a negatively charged substituent in this region in order to enhance the biological activity. The activity of compound **2** is larger than that of compound **15** because the former possesses a more electronegative substituent (acetoxy group) than **15** does (isobutyryloxy group) at this position. It is also observed that a blue contour occupies the C6 position of the B ring, implying that substituting an electronegative group at this position leads to a decrease in the biological activity, as shown by compound **18**, which has lower activity than compound **1** since it has a more electronegative group at this position.

Figure 7C depicts the hydrophobic contour map of the optimal model. The orange contour near the C9 position of the B ring indicates that a hydrophilic group at this position is essential for high activity, as exemplified by compound **4**, which is more potent and has a more hydrophilic group than compound **15** at the C9 position. Another orange contour at the C4 position of the A ring suggests that hydrophilic groups are favorable, as demonstrated by the fact that compound **2** is more active than **18** because **2** has a hydroxyl group at this position while **18** does not.

Insecticidal model

In the insecticidal model, the field applied in the CoMFA and CoMSIA analyses is the same as that used in the narcotic models. Four different models were also generated by adding Mlog*P* as an additional descriptor to the CoMFA and CoMSIA analyses. The results obtained from the four models are summarized in Table 4. For the 32 compounds, a cross-validated coefficient (R^2_{cv}) of 0.39 and a non-cross-validated coefficient (R^2_{ncv}) of 0.99 with five optimum components for the CoMFA model, an R^2_{cv} of 0.24 and an R^2_{ncv} of 0.99 with six ONCs for the CoMFA-Mlog*P* model, an R^2_{cv} of 0.46 and an R^2_{ncv} of 0.99 with six ONCs for the CoMSIA model, and an R^2_{cv} of 0.31 and an R^2_{ncv} of 0.98 with six ONCs for the CoMSIA-Mlog*P* model were observed. It was found that adding Mlog*P* to the CoMFA and CoMSIA analyses caused a reduction in R^2_{cv} and resulted in minor contributions (<10%) to both the CoMFA-Mlog*P* and CoMSIA-Mlog*P* models. The CoMSIA model was selected to be the optimal one based on its statistics (R^2_{cv}=0.46, R^2_{pred}=0.98). Graphs of the actual versus predicted pLD_{50} values of both the training and test sets using the optimal model are shown in Fig. 9.

CoMSIA contour maps

Figure 10A–C depicts the various contour maps of the optimal CoMSIA model. In the steric field, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80%- and 20%-level contributions, respectively. Similarly, in the electrostatic field, the blue (electropositive charge favorable) and red (electronegative charge favorable) contours represent 80%- and 20%-level contributions, respectively. In the CoMSIA hydrophobic

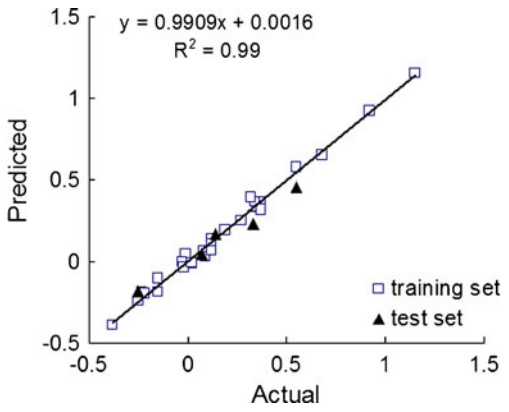

Fig. 9 Plot of actual versus predicted pLD_{50} values for the training and test compounds based on the optimal CoMSIA model for the insecticidal study

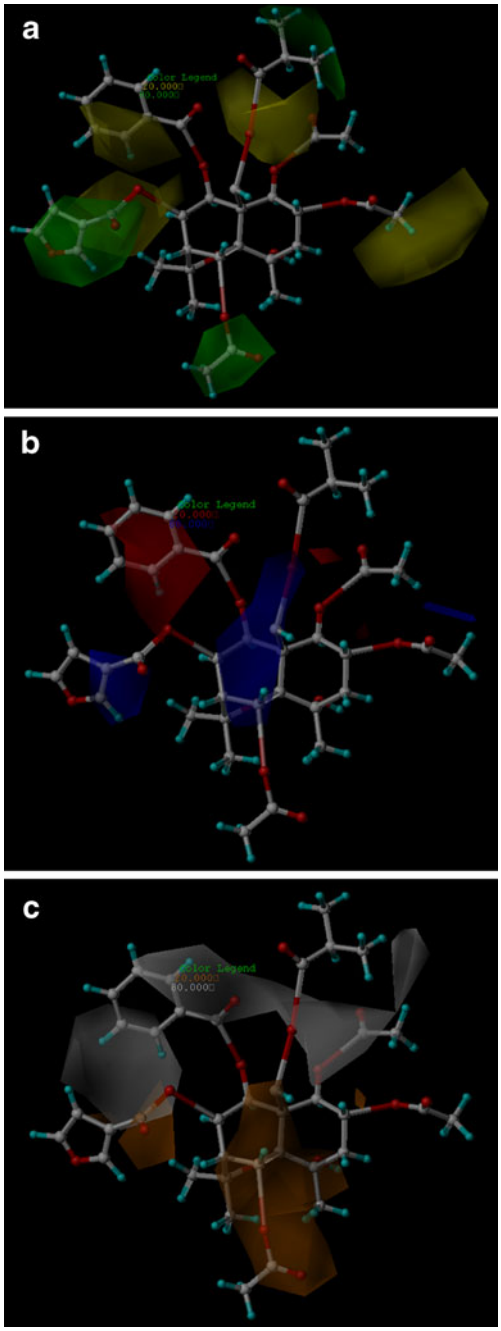

Fig. 10 CoMSIA stdev*coeff steric (**A**), electrostatic (**B**) and hydrophobic (**C**) contour maps for the insecticidal model. The color schemes of **A–C** are the same as in Fig. 7A–C. The most potent compound, **16**, is displayed as a reference

field, the white (hydrophobic favorable) and orange (hydrophobic unfavorable) contours represent 80%- and 20%-level contributions, respectively.

For the steric contours (Fig. 10A), the most active compound **16** is overlaid on the map. The green contour at the C6 position of the B ring indicates that a sterically bulky group is favored. This is further supported by the fact that compound **33** is more active than **22** because **33** has a more bulky substituent than **22** at this position. The yellow contour seen in the vicinity of the C2 substituent of the A ring indicates that the occupancy of this sterically unfavorable contour would have a detrimental effect on the biological activity. This correlation is verified by the many potent compounds (**16**, **25**, **27** and **40**) that have a less bulky substituent (acetoxy group) at this position. Less active compounds are also found to orient their substituents into this yellow contour map, as exemplified by the bulky substituent (isobutyryloxy group) in compound **12** and the 2-methylbutyryloxy group in compound **28**. The green contour near the C8 position of the B ring indicates that a bulky substituent here increases the biological activity. This is confirmed by compound **16**, which is more active than **34** as **16** has a bulky substituent at this position while **34** does not.

In Fig. 10B, the electrostatic contour map, compound **16** is again overlaid as the reference molecule. A large red contour is observed near the C9 position of the B ring, indicating the need for an electronegative group at this position to enhance biological activity. This is exemplified by compound **37**, which is more active than **34** because **37** possesses a more electronegative group than **34**. A blue region is found to surround the C8 position of the B ring, indicating that the decrease in bioactivity may be due to the presence of certain electronegative groups at this position. Compound **36**, which has no substituent, shows improved activity, while compound **13**, which has an electronegatively charged isobutyryloxy group at this position, is less active. A small negative charge disfavored blue region observed near the C2 position of the A ring suggests that an electronegative group at this position could reduce the biological activity, an observation that also correlates with experimental determinations. For example, compound **32** is less active than compound **38** because **32** has an electronegative group (acetoxy group) at this position while **38** does not.

Figure 10C displays the hydrophobic contour map of the optimal model with compound **16** overlaid. The orange and white contours highlight areas where hydrophilic and hydrophobic properties are preferred. A large orange contour map around the C6 position of the B ring indicates that its occupancy by hydrophilic groups would enhance the bioactivity. For most of the compounds, a hydroxyl or acetoxy group is present at this position, whereas for the weakest compound, **32**, no such substituent is present. The white contour located around the hydrophobic group at the C9 position of the B ring suggests that hydrophilic groups at this position would decrease biological activity. This is consistent with experimental results, as compound **39** is less active than **20**, **25** and **27** because **39** has a more hydrophilic (acetoxy) group than the corresponding benzoyloxy substituent in compounds **20**, **25** and **27**. The

orange contour falls near the substituent at the C8 position (B ring), suggesting an advantage of having hydrophilic substituents at this position in terms of activity, as demonstrated by compound **33**, which has higher activity than **36** because **33** has an acetoxy group at the C8 position while **36** does not.

Conclusions

A series of 43 natural β-dihydroagarofuran sesquiterpene polyesters were isolated from *Celastrus angulatus* and *Euonymus japonicus* and their insecticidal or narcotic activities against the fourth-instar larvae of *Mythimna separata* were evaluated. Our experimental results show that compounds **43** and **16** can be considered lead compounds for the development of potent insecticides with narcotic or insecticidal activities, respectively.

The structure–activity relationships of all of these molecules were also investigated by CoMFA and CoMSIA 3D-QSAR analyses, resulting in several optimal models with good activity prediction capabilities. In addition, by analyzing the optimal CoMSIA models, the observed variances in the insecticidal and narcotic activities can be explained. For the narcotic model, the electronic field is found to be the most influential among all three fields, contributing 52.2% to the optimal QSAR model. A similar result was also observed for the insecticidal model, where the electronic field again provided almost the greatest contribution (38.8%) to the best model. A hybrid effect of the electrostatic (38.8%) and hydrophobic (40.2%) interactions governs the insecticidal activities of the molecules. These may indicate that the electronic interaction plays the most important role in determining the biological activities of these natural compounds. In summary, the results from both the experimental and the theoretical investigations should prove valuable for the subsequent design of novel β-dihydroagarofuran sesquiterpene polyesters with enhanced activities.

Acknowledgments These projects were financed by the National Key S&T Research Foundation of China (2010CB126105) and the National Natural Science Foundation of China (30871663).

References

1. Grabley S, Thiericke R (1999) Adv Biochem Eng/Biotechnol 64:101–154
2. Cooper EL (2004) eCAM 1:215–217
3. Sparks TC, Crouse GD, Durst G (2001) Pest Manag Sci 57:896–905
4. Isman MB, Akhtar Y (2007) Plant natural products as a source for developing environmentally acceptable insecticides. In: Ishaaya I, Nauen R, Horowitz AR (eds) Insecticide design using advanced technologies. Springer, Berlin, pp 235–248
5. Gao JM, Wu WJ, Zhang JW (2007) Nat Prod Rep 24:1153–1189
6. Cortés-Selva F, Campillo M, Reyes CP, Jiménez IA, Castanys S, Bazzocchi IL, Pardo L, Gamarro F, Ravelo AG (2004) J Med Chem 47:576–587
7. Gonzalez AG, Jiminez IA, Ravelo AG, Coll J, Gonzalez JA, Lloria (1997) J Biochem Syst Ecol 25:513–519
8. Cheng CY, Huang PH (1999) Flora Reipublicae Popularis Sinicae 3(45). Science Press, Beijing, pp 7–128
9. Ji ZQ, Hu ZN, Liu GQ, Wu W (2004) J Acta Bot Boreali-Occidentalia Sin 24:748–753
10. Zhan JW, Wu WJ, Tian X (2004) Chin J Pesticide Sci 6:21–25
11. Wu W (1991) J Plant Prot 17:34–38
12. Spivey AC, Weston M, Woodhead S (2002) Chem Soc Rev 31:43–59
13. Cortes-Selva F, Jimenez IA, Munoz-Martinez F, Campillo M, Bazzocchi IL, Pardo L, Ravelo AG, Castanys S, Gamarro F (2005) Curr Pharm Des 11:3125–3139
14. Reyes CP, Muñoz-Martinez F, Torrecillas IR, Mendoza CR, Gamarro F, Bazzocchi IL, Núñez MJ, Pardo L, Castanys S, Campillo M, Jiménez IA (2007) J Med Chem 50:4808–4817
15. Zhang YL, Xu Y, Lin JF (1989) Acta Pharmacol Sin 24:568–578
16. Duan H, Takaishi Y, Momota H, Ohmoto Y, Taki T, Tori M, Takaoka S, Jia Y, Li D (2001) Tetrahedron 57:8413–8424
17. Kuo Y, King M, Chen C, Chen H, Chen C, Chen K, Lee KJ (1994) Nat Prod 57:263–269
18. González AG, San Andrés L, Ravelo AG, Luis JG, Jiménez IA, Domínguez XA (1989) J Nat Prod 52:1338–1341
19. Perez-Victoria JM, Tincusi BM, Jimenez IA, Bazzocchi IL, Gupta MP, Castanys S, Gamarro F, Ravelo AG (1999) J Med Chem 42:4388–4393
20. Cortes-Selva F, Munoz-Martinez F, Ilias A, Jimenez AI, Varadi A, Gamarro F, Castanys S (2005) Biochem Biophys Res Commun 329:502–507
21. Muñoz-Martinez F, Lu P, Cortes-Selva F, Perez-Victoria JM, Jimenez AI, Ravelo AG, Sharom FJ, Gamarro F, Castanys S (2004) Cancer Res 64:7130–7138
22. Ujita K, Takaishi Y, Tokuda H, Nishino H, Iwashima A, Fujita T (1993) Cancer Lett 68:129–133
23. Takaishi Y, Ujita K, Tokuda H, Nishino H, Iwashima A, Fujita T (1992) Cancer Lett 65:19–26
24. Wei SP, Wang MA, Zhang JW, Qian Y, Ji ZQ, Wu WJ (2009) Nat Prod Commun 4:461–466
25. Wei SP, Ji ZQ, Zhang JW (2009) Molecules 14:1396–1403
26. Wu WJ, Wang MA, Zhu JB, Zhou WM, Hu ZN, Ji ZQ (2001) J Nat Prod 64:364–367
27. Wu WJ, Tu YQ, Zhu JB (1992) J Nat Prod 55:1294–1298
28. Gonzalez AG, Gonzalez CM, Bazzochi IL, Ravelo AG, Luis JG, Dominguez XA (1987) Phytochemistry 26:2133–2135
29. Yamada K, Shizuri Y, Hirata Y (1978) Tetrahedron 34:1915–1920
30. Wu MJ, Zhao TZ, Shang Y (2004) J Chinese Chem Lett 15:41–42
31. Ji ZQ, Wu WJ, Yang H (2007) Nat Prod Res 21:334–342
32. Di Santo R, Fermeglia M, Ferrone M, Paneni MS, Costi R, Artico M, Roux A, Gabriele M, Tardif KD, Siddiqui A, Pricl S (2005) J Med Chem 48:6304–6314
33. Clark M, Cramer RD III, van Opdenbosch N (1989) J Comput Chem 10:982–1012
34. Nilsson J (1998) Multiway calibration in 3D QSAR. http://www.ub.rug.nl/eldoc/dis/science/j.nilsson
35. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130–4146

36. Viswanadhan VN, Ghose AK, Revenkar GR, Robins R (1989) J Chem Inf Comput Sci 29:163–172
37. Klebe G (1994) J Mol Biol 237:212–235
38. Staahle L, Wold S (1987) J Chemometr 1:185–196
39. Li YF, Liu YL, Song ZQ (2006) Agrochemicals 45:148–154
40. Roy KK, Dixit A, Saxena AK (2008) J Mol Graph Model 27:197–208
41. Golbraikh A, Tropsha A (2002) J Mol Graph Model 20:269–276
42. Clark RD, Sprous DG, Leonard JM (2001) Validating models based on large dataset. In: Holtje HD, Sippl W (eds) Rational approaches to drug design (Proceedings of the 13th European Symposium on Quantitative Structure–Activity Relationships). Prous Science, Barcelona, pp 475–485