ORIGINAL PAPER

Probing the structural requirements for antitubercular activity of scalarane derivatives using 2D-QSAR and CoMFA approaches

Suriyan Thengyai · Phornphimol Maitarat · Supa Hannongbua · Khanit Suwanborirux · Anuchit Plubrukarn

Abstract By use of 2D-QSAR and CoMFA approaches, a series of scalarane derivatives were assessed for which structural features confer antitubercular activity. The major contributions deduced from the 2D-QSAR models suggested positive effects on antitubercular activity of the shadow area on the XZ plane (representing the steric descriptor), the dipole moment, and relative positive charge (both representing the electrostatic descriptor). The contour map from the CoMFA model indicated the requirement for oxygenated functionality surrounding the furan moiety, particularly the extension of the acetoxy group either on C-16 toward the β plane or on C-19 toward the α plane. On the other hand, the region crowning C-12 was considered a sterically and negative-electrostatically disfavored area. Therefore, extension of a bulky negative functional group toward this area would diminish the binding interaction. The 2D-QSAR results agreed well with those from the CoMFA model and suggested the structural features of polarizable functionality that would enhance the antitubercular activity of the compounds in this series.

Keywords Antitubercular · CoMFA · Marine-derived sesterterpenes · 2D-QSAR · 3D-QSAR · Scalaranes

Introduction

Tuberculosis has long been one of the most notorious and deadliest infectious diseases that ever plagued human history. Despite tireless efforts from the WHO and from local governments worldwide, the disease has not been eradicated but has, instead, become more vigorous. This is especially the case when associated with the HIV pandemic, because it is one of the most encountered AIDS-related diseases. First-line drugs (i.e., isoniazid, rifampicin, pyrazinamide, ethambutol, and streptomycin) are completely effective when administered in combination, but this is an added difficulty in tuberculosis treatment. The standard regimens involving the above medicines, administered alone or in any recommended combinations, require a persistent 6–9-month-long treatment. Not surprisingly, such long-term treatment often leads to noncompliance, particularly among the poorly educated and the economically challenged. This eventually allows the bacteria to adapt and to develop into a deadlier and more problematic multi-drug-resistant strain.

We recently reported the isolation and modification of a series of marine-derived scalarane-type sesterterpenes, based on the parent structure of heteronemin (**1**) [1]. Compound **1** has long been known as a potent cytotoxin [2], and was also reported to possess antitubercular [3] and protein-function-inhibiting activity [4]. Over the past few

S. Thengyai · A. Plubrukarn (✉)
Marine Natural Products Research Unit (MNP),
Department of Pharmacognosy and Pharmaceutical Botany,
Faculty of Pharmaceutical Sciences, Prince of Songkla
University, Hat-Yai, Songkhla 90112, Thailand
e-mail: anuchit.pl@psu.ac.th

P. Maitarat · S. Hannongbua
Department of Chemistry, Faculty of Science,
Center of Nanotechnology, Kasetsart University,
Bangkok 10900, Thailand

K. Suwanborirux · A. Plubrukarn
Center of Bioactive Natural Products from Marine Organisms
and Endophytic Fungi (BNPME), Faculty of Pharmaceutical
Sciences, Chulalongkorn University, Patumwan,
Bangkok 10330, Thailand

years, other scalarane analogs were isolated and reported to be biologically active in various bioassay models, including antitubercular activity [5], cytotoxicity [6], and inhibiting activity against various enzymes [7]. Our first observation [2], consistent with previous findings by Crews and Bescansa [3], indicated that the oxygenated functionality surrounding the furan moiety were among the most crucial parts for affecting chemotherapeutically related activity. To extend such findings more systematically, two quantitative structure–activity relationship (QSAR) approaches, which included 2D-QSAR and 3D-QSAR (using comparative molecular field analysis, CoMFA, for the latter) [8], were used to investigate the antitubercular activity of scalaranes. The structural requirements for such activity, derived from the chemical structures of the available scalarane derivatives, have been extensively discussed (Scheme 1).

Results and discussion

Fourteen scalaranes from the dataset and their antitubercular activity are listed in Table 1. The univariate analysis of the compound dataset is summarized in Table 2. Despite the limited dataset size (with a marginal standard deviation of 0.86 on antitubercular activity) [9], the range of pMICs (3.05) and the resulting q^2 (see below) suggested that the dataset is statistically acceptable for both the 2D-QSAR and CoMFA analyses.

2D-QSAR models

Our first attempt to determine the SAR for antitubercular activity of the scalaranes was carried out using a 2D-QSAR approach. A total of 151 descriptors, calculated using Material Studio 4.3, were obtained for each compound in the dataset. For any descriptors that required molecular alignment, the structure of oxoheteronemin (**7**), which is the most active congener in the dataset and is therefore presumed to possess a conformation closest to the bioactive conformer, was arbitrarily adopted as a template (Fig. 1) [10]. The QSAR models were constructed using a stepwise
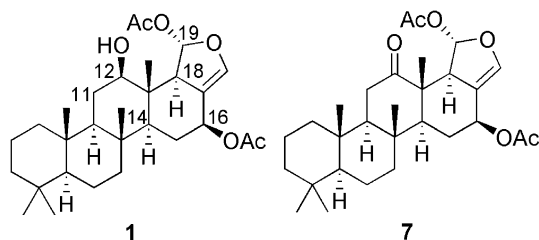
Scheme 1

procedure, from which three models (Eqs. 1–3) emerged. The calculated MICs taken from Eq. 3, because of the best resulting q^2, were plotted against the observed values (Fig. 2a) to demonstrate the good predictability of the model. The plot of residual MICs (Fig. 3), also determined using Eq. 3, shows only a slight deviation between the observed and the calculated MICs of each compound in the dataset. The exception was for **10**, the calculated MIC of which was distorted as badly as −254 μM from the observed value.

$$pMIC = 0.110\,(\pm 0.027)\,SHDW_{XZ} - 4.348\,(\pm 2.225)$$
$$r^2 = 0.581;\ q^2 = 0.423;\ SEE = 0.580;\ F = 16.61$$

$$(1)$$

$$pMIC = 0.098\,(\pm 0.020)\,SHDW_{XZ}$$
$$+\ 0.429\,(\pm 0.123)\,\mu - 4.976\,(\pm 1.610)$$
$$r^2 = 0.801;\ q^2 = 0.643;\ SEE = 0.417;\ F = 22.17 \quad (2)$$

$$pMIC = 0.116\,(\pm 0.018)\,SHDW_{XZ} + 0.434\,(\pm 0.103)$$
$$\mu + 12.070\,(\pm 5.127)\,RPCG - 7.937\,(\pm 1.849)$$
$$r^2 = 0.872;\ q^2 = 0.689;\ SEE = 0.351;\ F = 22.73 \quad (3)$$

where $SHDW_{XZ}$ is the shadow of the structure projecting on XZ plane; μ is the total dipole moment of the whole molecule; and RPCG is the positive charge of the most positive atom relative to the total positive charge. (Note: the numbers in parentheses are the standard deviations for the coefficients.)

Despite barely mediocre r^2 and q^2, Eq. (1) indicated the primary and most important descriptor, the shadow area on the XZ plane ($SHDW_{XZ}$). The predictability of the models was improved significantly in Eqs. (2) and (3), in which two other electrostatic descriptors (the total dipole moment, μ, and the relative positive charge, RPCG) were added. All three descriptors contributed positive coefficients to the inhibitory concentration against mycobacterium (Tables 3, 4).

Generally, the shadow area is a spatial descriptor, and therefore implies the steric effect on activity. The XZ plane discussed here was defined as a plane perpendicular to the carbocyclic plane and across from left to right of the structure when drawn as shown. The positive coefficients of $SHDW_{XZ}$ observed in all three equations suggested that a bulky substituted functionality extending along the X and/ or Z axes could exert its positive effect on antitubercular activity. This can be seen in **1**, **2**, and **7** ($SHDW_{XZ} =$ 90.453, 92.245, and 88.979 Å^2), in which the extension of the 16-OAc group provided a positive effect of the shadow area along the X axis. As for the shadow area along the Z axis, a similar effect was observed with α-substitution on either C-19 or C-18, as seen in **5** and **12** ($SHDW_{XZ} =$ 85.727 and 82.775 Å^2). On the other hand, the compact structural feature of **8** and **10** ($SHDW_{XZ} =$ 78.896 and
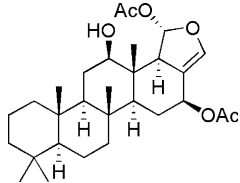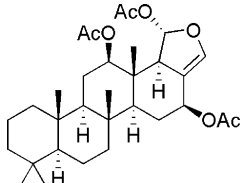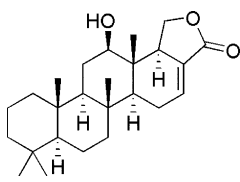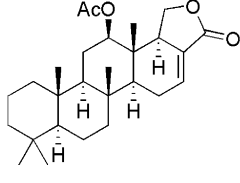
Table 1 Compound data set
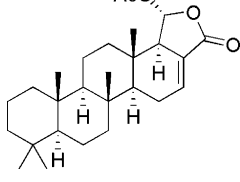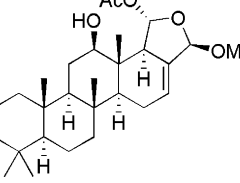
Entry		Observed		Calculated			
		pMIC	MIC (μM)	2D-QSAR		CoMFA	
				pMIC	MIC (μM)	pMIC	MIC (μM)
1		5.52	3	5.30	5.01	5.50	3.16
2		5.52	3	5.82	1.51	5.63	2.34
3		4.80	16	5.10	7.94	4.65	22.39
4		3.93	117	4.35	44.67	3.95	112.20
5		5.40	4	5.14	7.24	5.37	4.27
6		4.27	54	4.76	17.38	4.30	50.12

Table 1 continued

Entry		Observed		Calculated			
				2D-QSAR		CoMFA	
		pMIC	MIC (μM)	pMIC	MIC (μM)	pMIC	MIC (μM)
7		6.64	0.23	6.26	0.55	6.57	0.27
8		3.59	257	3.65	223.87	3.53	295.12
9		4.85	14	4.60	25.12	4.83	14.79
10		3.87	135	3.41	389.05	3.89	128.82
11		4.19	64	4.18	66.07	4.34	45.71
12		5.10	8	5.03	9.33	5.20	6.31

Table 1 continued

Entry		Observed		Calculated			
				2D-QSAR		CoMFA	
		pMIC	MIC (μM)	pMIC	MIC (μM)	pMIC	MIC (μM)
13		3.89	130	4.14	72.44	3.96	109.65
14		4.24	58	4.06	87.10	4.10	79.43
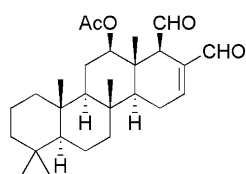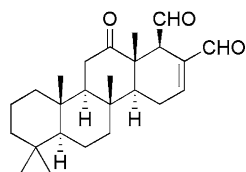
Table 2 Univariate analysis of the compound dataset

Univariate analysis	
Number of sample point, n	14
Range of pMIC	3.05
Maximum pMIC	6.64
Minimum pMIC	3.59
Mean, $\bar{x}$	4.70
Variance, s^2	0.69
Standard deviation, s	0.86

Fig. 1 Aligned structures of compounds in the dataset, using **7** as structural template

73.650 $\mathring{A}^2$) resulted in reduced activity of the two compounds.

The electrostatic descriptors (i.e., the dipole moment, μ, and the relative positive charge, RPCG) exerted their effect in such a way that the structure required functional groups that could induce the positive electrostatic field, either globally throughout the structure or locally over any given atoms. The least polarizable analogs, e.g., **8** and **10** ($\mu = 2.625$ and 3.032), are negatively affected and therefore offered lower potency. Other highly polarizable scalaranes, e.g., **2**, **7**, and **12** ($\mu = 4.441$, 5.992, and 4.542), are otherwise more active. Note that for compounds with contrasting steric and electrostatic descriptors (i.e., those with larger SHDW$_{XZ}$ and smaller μ and/or RPCG, and vice versa), the steric contribution seemed to prevail. Examples of this included **12** and **13**, both of which possess a strong electrostatic field in a close range ($\mu = 4.542$ and 4.769, and RPCG $= 0.115$ and 0.086). However, the expanded structure of **12** (SHDW$_{XZ} = 82.775$ $\mathring{A}^2$) led to considerably

more potent activity than that of **13** (SHDW$_{XZ} = 77.347$ $\mathring{A}^2$).

However, the requirements of the electrostatic descriptors, μ and RPCG, are paradoxical and surprising. For most antitubercular compounds, it is generally believed that they require effective lipophilicity to penetrate the thick layer of the mycolic acid-enriched membrane of the mycobacteria. The presence of μ and RPCG in Eqs. (2) and (3) could therefore be controversial. Nevertheless, the large carbocyclic moiety should be lipophilic enough to carry the scalarane molecule efficiently into bacterial cells ($c\log P$ of each compound in dataset is shown in Table 3). Furthermore, the effects from both electrostatic descriptors may simply indicate the requirement for the polarizable

Fig. 2 Representative plots between observed and calculated pMICs from (a) 2D-QSAR (expressed as $y = 0.87x + 0.59$; $r^2 = 0.8714$); and (b) the CoMFA approach (expressed as $y = 0.99x + 0.05$; $r^2 = 0.9895$)
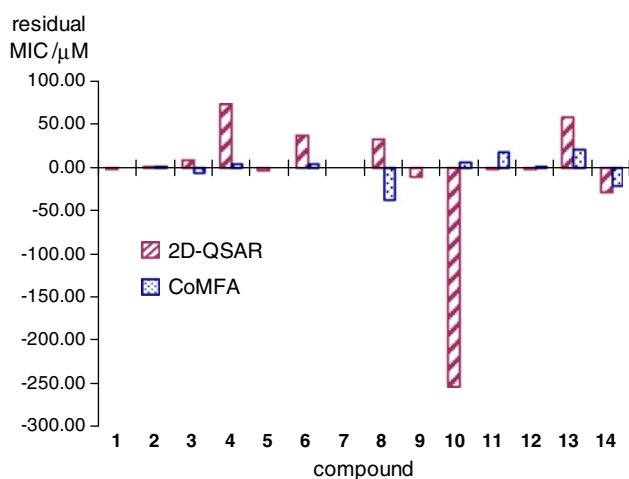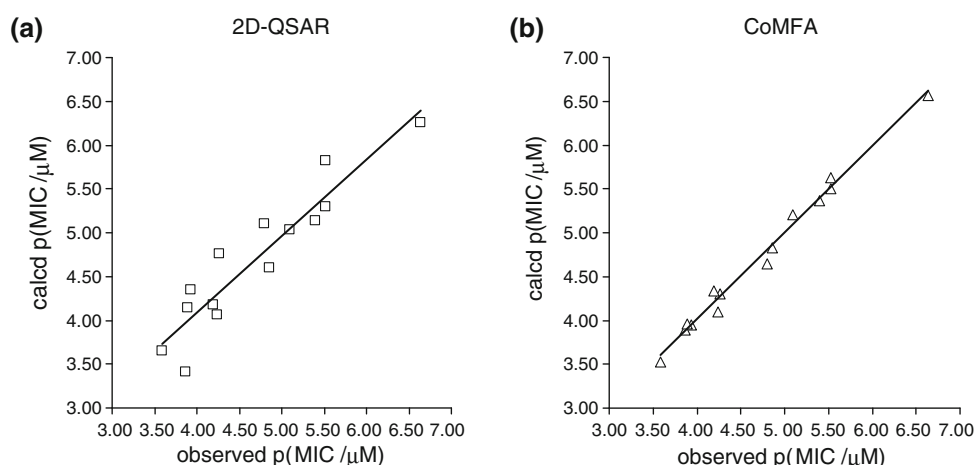

Fig. 3 Residual MICs (actual MIC–predicted MIC) of each compound in the dataset, compared between the MICs from the 2D-QSAR and CoMFA approaches

Table 3 Calculated molecular descriptors (SHDW$_{XZ}$, μ, RPCG) and $c \log P$ of compounds in the dataset

Compound	SHDW$_{XZ}$ ($\mathring{A}^2$)	μ	RPCG	$c \log P$
1	90.453	3.545	0.100	4.454
2	92.245	4.441	0.094	5.204
3	76.916	5.009	0.161	5.054
4	77.293	3.611	0.146	5.183
5	85.727	3.231	0.144	5.891
6	87.986	2.800	0.106	5.163
7	88.979	5.992	0.106	4.982
8	78.896	2.625	0.108	4.861
9	82.816	3.448	0.119	5.213
10	73.650	3.032	0.124	5.737
11	77.625	3.965	0.115	4.007
12	82.775	4.542	0.115	4.007
13	77.347	4.769	0.086	4.534
14	77.892	3.253	0.128	4.136

Table 4 Pearson correlation matrix of selected descriptors

	SHDW$_{XZ}$	μ	RPCG
SHDW$_{XZ}$	1.000	–	–
μ	0.182	1.000	–
RPCG	−0.468	−0.066	1.000

pharmacophores, presumably to interact with the hydrophilic binding sites.

CoMFA model

In addition to a rather abstract mathematical prediction from the 2D-QSAR models, a more pictorial analysis of the QSAR of scalaranes based on a CoMFA approach was performed. All 14 compounds were used as the training set in this study. Based on a similar argument given for the 2D-QSAR analysis, the structure of **7** was again arbitrarily chosen as a template for the structural alignment. A series of alignment rules, including fit-atom, matching, and superimposition, were attempted. However, no significant differences were observed between the resulting models. This was possibly because of the highly rigid nature of the tetracarbocyclic skeleton that fixed the aligned structures in such a manner that any deviating factors (caused by a

different methodology) did not significantly affect the alignments. In this study hereafter, the superimposition alignment was conveniently chosen.

The attempted alignments based on various grid spaces (1, 2, and 3 Å) and probe atoms (sp^3 C, sp^3 O, and H) led to the most suitable model when sp^3 C was probed within a 2-Å spacing ($r^2 = 0.988$; $q^2 = 0.668$; $F = 130.569$; SEE = 0.121). The plot between calculated and observed pMICs (Fig. 2b) is even more linear than that obtained

Table 5 Attempted trails in CoMFA experiments, based on superimposition alignment

Grid space (Å)	Probed atom[a]	Statistical data					Relative contribution	
		q^2	r^2	SEE	OPC[b]	F value	Steric	Electrostatic
1	sp^3 C (+1)	0.605	0.997	0.068	6	348.63	0.449	0.551
	sp^3 O (−1)	0.533	0.995	0.079	6	254.47	0.382	0.618
	H (+1)	0.527	0.974	0.166	4	85.21	0.389	0.611
2	sp^3 C (+1)	0.668	0.988	0.121	5	130.57	0.383	0.617
	sp^3 O (−1)	0.589	0.995	0.085	6	220.47	0.355	0.645
	H (+1)	0.545	0.964	0.195	4	60.97	0.383	0.617
3	sp^3 C (+1)	0.573	0.959	0.208	4	53.09	0.358	0.642
	sp^3 O (−1)	0.717	0.984	0.138	5	99.62	0.316	0.684
	H (+1)	0.810	0.988	0.119	5	133.63	0.311	0.689

[a] The numbers in parentheses indicate the charge of each probed atom

[b] Optimum number of components

from the 2D-QSAR approach. The calculated MICs, with a small deviation range of −38 to +20 μM compared with that of −254 to +72 μM for the 2D-QSAR (Fig. 3), were more precisely predicted. It should be mentioned that when the grid space of 3 Å was attempted with H as the probed atom, the model yielded a better predictability with a q^2 of 0.810. However, the model was dismissed as the attempted grid size was considered too large for the tiny probed atoms (Table 5).

The results from the selected CoMFA model provided a ratio of 0.38:0.62 between steric and electrostatic contributions. The rather high proportion of electrostatic contribution reflected the observation regarding μ and RPCG descriptors made earlier in our 2D-QSAR models. This also suggested that the polarizable functionality were required to enhance antitubercular activity. However, it should be considered that the dataset was dominated by the oxygenated functionalized scalaranes and could therefore result in a high ratio of electrostatic contribution.

The contour map of the ideal binding sites based on the structure of **7** (Fig. 4) agreed well with the observations made earlier by Crews and Bescansa [3] and by us [2] that the functionality affecting the antitubercular activity of the scalaranes clustered mainly in close proximity to the furan moiety, especially on the "eastern" side extending from C-16, and on the "northern" side hovering over C-19 and C-12. The positive effect of the acetoxy functionality on C-16 was highly pronounced in our CoMFA model and confirmed our speculation regarding the effect of the SHDW$_{XZ}$ descriptor discussed earlier for Eqs. 1–3. The green (sterically favored) contour on the far end from C-16 indicated the steric field favored by the acetoxy group (e.g., **1**, **2**, and **7**), whereas the nearby red (positive-electrostatically disfavored) region suggested a restricted orientation in such a way that the carbonyl oxygen should favorably turn toward the α plane. In addition, another
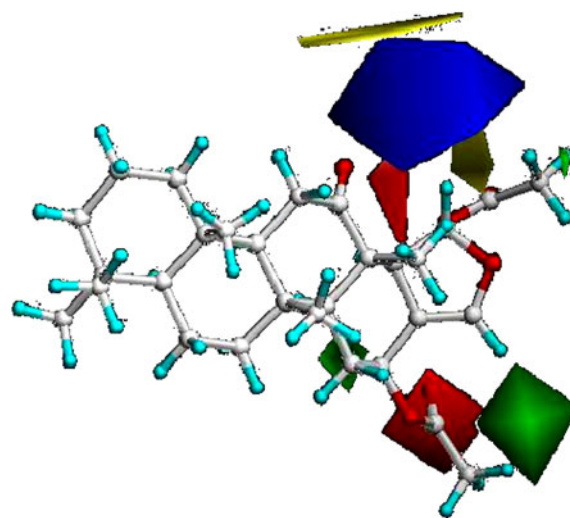

Fig. 4 Contour map of ideal binding sites for antitubercular scalaranes, based on the structure of **7**. Color code for each contour is as follows; *blue* positive-electrostatically favored, *red* positive-electrostatically disfavored, *green* sterically favored, and *yellow* sterically disfavored

small green contour underneath C-16 reflected the reduced activity of all the Δ^{15} and Δ^{16} analogs, including **4**, **6**, **8**, **10**, **11**, **13**, and **14**. Even in the cases of **5** and **12**, both of which expressed strong potency (that could be attributed to the extension from C-19 and C-18, respectively), their activity was slightly less potent than that of other active analogs. This was presumably because of the Δ^{16} olefinic functionality.

The effects of the 19-α-OAc group (e.g., as in **1**, **2**, **5**, and **7**), suggested by the green and red patches that extend toward the uppermost end of the structure, could be discussed in the same manner as those for 16-β-OAc. More interestingly, hovering over C-12 and slightly toward the left-hand side of the furan moiety were the yellow

(sterically disfavored) and blue (positive-electrostatically favored) contours. This suggested that any large and negatively polarized (and presumably oxygenated) functionality extending from C-12 could hamper the activity of the compounds. In addition, although there was only one example in the dataset (**12**), the functionality protruding down the α-plane from C-18 (i.e., C-19 aldehyde of **12**) also showed a sterically favored result. This was due to an orientation similar to that of the 19-α-OAc group.

Conclusion

The structure–activity relationships of 14 scalarane derivatives were assessed using 2D-QSAR and CoMFA approaches. Both models coherently demonstrated the requirement for oxygenated functionality surrounding the furan moiety. The 16-OAc and 19-OAc, regarded as steric and negative-electrostatic functionality, were found to be strongly important for antitubercular activity. On the other hand, the extended negatively polarized functional groups crowning C-12 disfavored potency, suggesting the non-steric, positively electrostatic requirement in the vicinity of this region. The strong electrostatic contributions observed in both 2D-QSAR and CoMFA models, which contradicted the general belief that antitubercular drugs should be lipophilic, may merely reflect the limitation of the dataset, as all the compounds in the dataset possess a similar rigid tetracarbocyclic skeleton that is characteristic of scalaranes. A better generalized model based on a larger variety of scalaranes could offer insight on which structural modifications would improve antitubercular and other chemotherapeutically related activities. To devise such models, more extensive derivatization of the scalaranes are now underway.

Methods

Compound dataset

The dataset comprised 14 sacalarane derivatives (Table 1). The antitubercular activity of each compound, expressed as MICs, was taken directly from our previous reports [1]. All calculations were performed based on pMIC converted from the MIC of each compound. Biological activity was determined on the basis of the microplate alamar blue assay [1, 11] (Central Bioassay Lab, Biotec, Thailand) to ensure that all were carried out in the same laboratory, and therefore can be confidently compared. All the compounds in the dataset were used without elimination, presuming that each was equally crucial in constructing the QSAR models.

Computational studies

2D-QSAR

The three-dimensional structure for each compound in the dataset was initially optimized, and the energy refinements were carried out based on COMPASS force field and Gasteiger charge calculations using Material Studio 4.3 [12]. The structure of **7** was used as a template when a molecular alignment was required for the calculations of any comparative descriptors. All alignments were executed in a manner that the carbocyclic moiety (C-1–C-14) was fixed as shown in Fig. 1. All the calculated descriptors were filtered to eliminate those with a value of 0. This led to 151 usable descriptors, which were subjected to further computational studies.

The construction of QSAR models was carried out on the basis of multiple linear regression (SPSS 12) with a stepwise procedure for the selection of highly correlated descriptors. The predictive ability of the developed models was evaluated on the basis of the cross-validated partial least-square leave-one-out procedure and was expressed as q^2. The results were also represented by the conventional non-cross-validated correlation coefficient, r^2; standard error of estimate, SEE; and the ratio of variance, F value. Details of the calculations can be found elsewhere [13].

CoMFA

For the CoMFA experiment, the three-dimensional structure of each compound in the dataset was preoptimized with MM+ and was energetically minimized with TRIPOS molecular mechanics (SYBYL 7.0) [14]. Gasteiger–Huckel charge was loaded on each atom of each molecule. Each compound in the training set was individually aligned on **7** as a structural template using fit-atom, matching, and superimposition as attempted alignment rules. The trials for the alignment also included grid spaces of 1, 2, and 3 Å, in each of which the probed atoms were sp^3 C (+1), sp^3 O (−1), and H (+1). The electrostatic and steric potential energies were separately calculated between grid probes and molecules of interest (based on coulombic and Lennard–Jones equations). The default repulsion cut-off for molecular field calculation was 125 kJ/mol, with the minimal sigma of 2 Å [18]. Similar to the QSAR model, the regression analysis was carried out using the cross-validated partial least-square leave-one-out method, and expressed as q^2, with the results of the model represented by r^2, SEE, and F value [13].

References

1. Wonganuchitmeta S, Yuenyongsawad S, Keawpradub N, Plubrukarn A (2004) J Nat Prod 67:1767
2. Jaisamut S, Thengyai S, Yuenyongsawad S, Karalai C, Plubrukarn A, Suwanborirux K (2009) Pure Appl Chem 81:1019
3. Crews P, Bescansa P (1986) J Nat Prod 49:1041
4. Doi Y, Shigamori H, Ishibashi M, Mozobe F, Kawashima A, Nakaike S, Kobayashi J (1993) Chem Pharm Bull 41:2190
5. Kobayashi M, Okamoto T, Hayashi K, Yokoyama N, Sasaki T, Kitagawa I (1994) Chem Pharm Bull 42:265
6. Ryu G, Matsunaga S, Fusetani N (1996) J Nat Prod 59:515
7. Fontana A, Cavaliere P, Ungur N, D'Souza L, Parameswaram PS, Cimino G (1999) J Nat Prod 62:1367
8. El Sayed KA, Bartyzel P, Shen X, Perry TL, Zjawiony JK, Hamann MT (2000) Tetrahedron 56:949
9. Ledroit V, Debitus C, Ausseil F, Raux R, Menou J, Hill BT (2004) Pharm Biol 42:454
10. Youssef DTA, Shaala LA, Emara S (2005) J Nat Prod 68:1782
11. Pettit GR, Tan R, Cichacz ZA (2005) J Nat Prod 68:1253
12. Hernández-Guerrero CJ, Zubía E, Ortega MJ, Carballo JL (2006) Tetrahedron 62:5392
13. Kamel HN, Kim YB, Rimoldi JM, Fronczek FR, Ferreira D, Slattery M (2009) J Nat Prod 72:1492
14. Mahidol C, Prawat H, Sangpetsiripan S, Ruchirawat S (2009) J Nat Prod 72:1870
15. Nam S, Ko H, Shin M, Ham J, Chin J, Kim Y, Kim H, Shin K, Choi H, Kang H (2006) Bioorg Med Chem Lett 16:5398
16. Nam S, Ko H, Ju M, Hwang H, Chin J, Ham J, Lee B, Lee J, Won D, Choi H, Ko J, Shin K, Oh T, Kim S, Rho J, Kang H (2007) J Nat Prod 70:1691
17. Williams DE, Hollander I, Feldberg L, Frommer E, Mallon R, Tahir A, van Soest R, Anderson RJ (2009) J Nat Prod 72:1106
18. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959
19. Hansch C, Leo A (1995) Exploring QSAR fundamentals and applications in chemistry and biology. American Chemical Society, Washington DC
20. Gedeck P, Rohde B, Bartels C (2006) J Chem Inf Model 46:1924
21. Tebib S, Bourquiqnon J, Wermuth CG (1987) J Comput Aided Mol Des 2:153
22. Collins LA, Franzblau SG (1997) Antimicrob Agents Chemother 41:1004
23. Material Studio 4.3. Accelrys Software Inc., San Diego, CA
24. Maitarad P, Kamshonwongpaisan S, Vanichtanankul J, Vilaivan T, Yuthavong Y, Hannongbua S (2009) J Comput Aided Mol Des 23:241
25. SYBYL 7.0. TRIPOS Associates Inc., St. Louis, MO