Original article

# 2D- and 3D-QSAR studies on 54 anti-tumor Rubiaceae-type cyclopeptides

He Yan [a,b], Xulin Pan [a,*], Ninghua Tan [a,*], Junting Fan [a,b], Guangzhi Zeng [a], Hongjin Han [a]

[a] State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650204, China
[b] Graduate School of the Chinese Academy of Sciences, Beijing 100039, China

## ARTICLE INFO

## ABSTRACT

RA-VII, a bicyclic hexapeptide isolated from the roots of *Rubia cordifolia*, *Rubia akane* belongs to Rubiaceae-type cyclopeptides (RAs) and has attracted much attention for its potent anti-tumor activity and its bicyclic structure incorporating the isodityrosine moiety. In this work, hologram quantitative structure–activity relationship (HQSAR), comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA) methods were employed to develop 2D- and 3D-QSAR models for 54 anti-tumor RAs. The LOO cross-validated $q^2$ values of HQSAR, CoMFA and CoMSIA models are 0.701, 0.510 and 0.613, respectively. The predictive ability of these models was validated by the test set including 7 RAs, and the predicted $IC_{50}$ values were in good agreement with the experimental $IC_{50}$ values. HQSAR result showed that chirality descriptor plays an important role in anti-tumor activity of RAs and OMe at $R_1$ and $R_2$ is necessary for increasing their activity. CoMFA and CoMSIA results demonstrated that small bulky and electropositive side chains at $R_3$ position and hydrophobic groups at $R_7$ and $R_8$ positions will increase their activity, and intra-molecular hydrogen bonds between residues 1 and 4 are necessary to maintain the pharmacophoric conformation of RAs. These results may be helpful in designing novel and potential anti-tumor RAs.

## 1. Introduction

Rubiaceae-type cyclopeptides (RAs) [1,2] are homodicyclohexapeptides mainly formed with one D-α-alanine, one L-α-alanine, three modified N-methyl L-α-tyrosines and one other protein L-α-amino acid (Fig. 1). The most unusual feature is a 14-membered ring formed by oxidative coupling of the phenolic oxygen of one tyrosine with a carbon ortho to the phenolic hydroxyl group of an adjacent tyrosine with a *cis* peptide bond and the 14-membered ring was fused to the 18-membered cyclic hexapeptide ring. RAs were originally isolated from *Bouvardia ternifolia* (Rubiaceae), later from *Rubia cordifolia*, *Rubia akane* and *Rubia yunnanensis* (Rubiaceae). Until now, 23 RAs have been isolated from higher plants. RAs possess a promising anti-tumor activity *in vitro* and *in vivo*, and the major active principle RA-VII (**44**) was reported to be underwent phase I clinical trials at NCI as an anti-cancer drug in Japan in 1990s, and its therapeutic ratio was 400. It has been reported that RAs mainly interacted with 80S ribosomes and inhibited protein synthesis. Moreover, RAs changed the conformational structure of actin to cause $G_2$ arrest by the inhibition of cytokinesis. The ^{1}H NMR

spectrum of RA-VII suggested the presence of two stable conformational states in CDCl₃, conformers A and B, and of three different conformers A, B and C in a polar solvent, such as DMSO-$d_6$. The predominant conformer A exhibited a *trans* peptide bond at L-Ala[2] and L-Tyr[3] by stabilization of the intra-molecular hydrogen bond between D-Ala[1]–CO and L-Ala[4]–NH, while conformer B exhibited a *cis* peptide bond at L-Ala[2] and L-Tyr[3], and conformer C adopted three *cis* peptide bonds at L-Ala[2] & L-Tyr[3], L-Ala[4] & L-Tyr[5], L-Tyr[5] & L-Tyr[6]. Among these conformers, a *cis* peptide bond at L-Ala[2] and L-Tyr[3] plays an important role in RAs' anti-tumor activities.

As part of an effort to develop novel anti-tumor cyclopeptides from plants, researchers from our laboratory have recently isolated some anti-tumor RAs from *R. yunnanensis* (unpublished). Moreover, in order to study the anti-tumor mechanism of RAs, reverse docking method [3] was employed to find potential targets of RAs and several targets were selected to validate by experimental methods (unpublished). Presently there is no quantitative structure–activity relationships' (QSAR) research of RAs, which would be greatly helpful in illustrating the mechanism of anti-tumor activity of RAs. In the present study, Hologram quantitative structure–activity relationship (HQSAR), comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA) [4–7] have been employed to study anti-tumor activity of 54 RAs published by Itokawa et al. [8–16]. This analysis provides

 * Corresponding authors. Tel./fax: +86 871 5223800.
   E-mail addresses: panxl@mail.kib.ac.cn (X. Pan), nhtan@mail.kib.ac.cn (N. Tan).
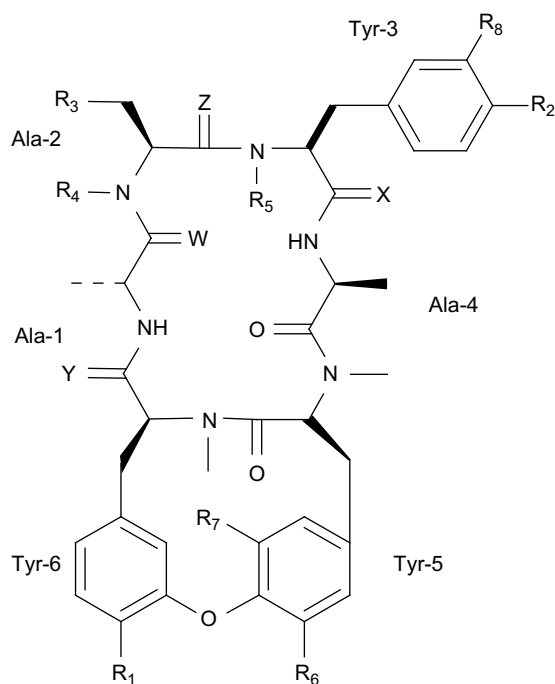
**Fig. 1.** Structural skeleton of RAs.

a platform for the prediction of novel RAs anti-tumor activities and subsequently enables their interpretation of anti-tumor mechanism.

## 2. Computational

### 2.1. Data sets

The training and test sets comprise 54 RAs published by Itokawa et al. [8–16], which have been shown to possess cytotoxicity against lymphocytic leukaemia P388 cell. The biological values covering a wide range from nanomolar to micromolar were converted into the corresponding $pIC_{50}$ ($-\log IC_{50}$) values and used as dependent variables in the HQSAR, CoMFA and CoMSIA analyses. According to the substitutes at positions $R_1$–$R_8$, X, Y, Z and W (Fig. 1), 54 RAs were randomly divided into the training set (47 compounds) and test set (7 compounds, labeled with an asterisk). The test compounds were selected manually considering the structural diversity and wide range of activity in the data set. Chemical structures and corresponding biological data are shown in Table 1.

### 2.2. Molecular modeling

The three-dimensional structures were constructed by change in the crystal structure of [Tyr-3–ψ(CH₂NH)–Ala-4]RA-VII (Fig. 2, three *cis* peptide bonds at L-Ala² and L-Tyr³, L-Ala⁴ and L-Tyr⁵, L-Tyr⁵ and L-Tyr⁶) [16] using SYBYL 6.9 software [4,17]. Optimized structures were performed by using PM6 method which was implemented in MOPAC 7.0 (Fujitsu) software [18,19] and then the Gasteiger–Huckel charge was added by using SYBYL 6.9.

### 2.3. Molecular alignment for 3D-QSAR analysis

The common cyclopeptide ring of RAs is rigid, this was chosen to align all molecules, while the most active compound **45** [13] was used as the template for superimposition (Fig. 3). Each analog was aligned to the template by rotation and translation so as to minimize the RMSD between atoms in the template and the corresponding atoms in the analog using the DATABASE ALIGN option in SYBYL. The aligned compounds are displayed in Fig. 4.

### 2.4. Calculation of HQSAR descriptors

HQSAR is a novel QSAR method that eliminates the need for 3D structure, putative binding conformations, and molecular alignment [20]. In HQSAR, each molecule in the data set is divided into all possible structural fragments (linear, branched, and overlapping) containing user defined minimum and maximum number of atoms. Each unique fragment is assigned a specific large integer that corresponds to a bin in an integer array of fixed length L (L is generally in the range 50–500). Standard PLS analysis is then applied and leave-one-out cross-validation is applied to determine the number of components that yield optimally predictive model. During fragment generation, identical fragments are always hashed to the same bin, and the corresponding occupancy for that bin is incremented. However, as the hologram length is generally smaller than the total number of unique fragments, different unique fragments can hash to the same bin, and this phenomenon is called 'fragment collision' [18]. In order to reduce the probability of identical or similar fragment collision occurring, all 21 HQSAR analyses in this work were performed by set values of hologram length with all the 153 odd numbers from 97 to 401. Moreover, HQSAR models can be affected by a number of parameters concerning hologram generation: fragment size, fragment distinction, and so on. Several combinations of these parameters were considered during the HQSAR modeling runs.

### 2.5. Calculation of CoMFA and CoMSIA descriptors

The initial CoMFA and CoMSIA models were calculated using the SYBYL 6.9 molecular modeling software [5].
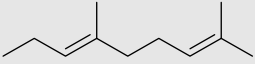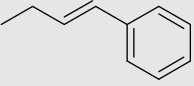
For CoMFA calculations, steric and electrostatic interactions were calculated using an sp³ carbon atom and a +1 charge as steric and electrostatic probes, respectively, and Tripos force field with a distance-dependent dielectric constant at all intersections in a regularly spaced grid (2 Å). The default value of 30 kcal/mol was set as the maximum steric and electrostatic energy cutoff. The minimum column filtering was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold. Regression analysis was performed using the cross-validation of compounds leave-one-out (LOO) method [5]. The non-cross-validated conventional analysis was produced with the optimal number of components equal to that yielding the highest $q^2$, and the corresponding conventional correlation coefficient $r^2$, its standard error, and the F ratio were also calculated.

Five CoMSIA similarity index fields (steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor) were evaluated using the sp³ carbon probe atom with a radius of 1 Å and a +1 charge placed at the lattice points of the same region of grid as it was used for the CoMFA calculations. A distance-dependent Gaussian type was used between the grid point and each atom of the molecule. The default value of 0.3 was used as the attenuation factor. The minimum column filtering was set to 1.0 kcal/mol. The statistical evaluation for the CoMSIA analysis was carried out in the same way as described in CoMFA.

## 3. Results and discussion

Results of HQSAR, CoMFA, and CoMSIA models are shown in Table 4. Experimental and predicted $pIC_{50}$ values of the training and test set (labeled with an asterisk) compounds are shown in Table 5.

**Table 1**
Structures and experimental pIC50 values of the training and test set (labeled with an asterisk) compounds.
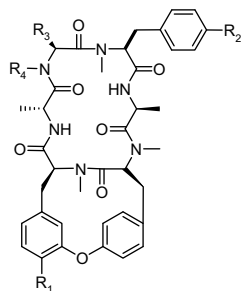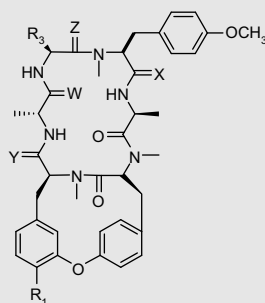


| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $pIC_{50}$ |
|-----|-------|-------|-------|-------|------------|
| 1[*] | $OCH_3$ | $OCH_3$ | $CH_2CH_2COOH$ | H | 6.66 |
| 2 | $OCH_3$ | $OCH_3$ | $(CH_2)_2COOCH_3$ | H | 7.23 |
| 3 | $OCH_3$ | $OCH_3$ | $CH_2COOCH_3$ | H | 7.38 |
| 4 | $OCH_3$ | $OCH_3$ | $CH_2OH$ | H | 7.85 |
| 5[*] | $OCH_3$ | $OCH_3$ | $CH_2SAc$ | H | 8.01 |
| 6 | $OCH_3$ | $OCH_3$ | $(CH_2S-)_2$ | H | 6.20 |
| 7 | $OCH_3$ | $OCH_3$ | $CH_2SH$ | H | 6.44 |
| 8 | $OCH_3$ | $OCH_3$ | $(CH_2)_3OH$ | H | 6.76 |
| 9 | $OCH_3$ | $OCH_3$ | $CH_2CH=CH_2$ | H | 7.41 |
| 10 | $OCH_3$ | $OCH_3$ | $CH_2CHO$ | H | 7.47 |
| 11 | $OCH_3$ | $OCH_3$ | $(CH_2)_4OH$ | H | 6.66 |
| 12 | $OCH_3$ | $OCH_3$ | $(CH_2)_2CH_3$ | H | 7.86 |
| 13 | $OCH_3$ | $OCH_3$ | $(CH_2)_3N_3$ | H | 7.45 |
| 14 | $OCH_3$ | $OCH_3$ | $CH_2COOH$ | H | 5.92 |
| 15 | $OCH_3$ | $OCH_3$ | $(CH_2)_2OH$ | H | 7.22 |
| 16 | $OCH_3$ | $OCH_3$ | $CH=CH_2$ | H | 7.59 |
| 17 | $OCH_3$ | $OCH_3$ | $(CH_2)_2SCH_3$ | H | 7.69 |
| 18[*] | H | H | $CH_3$ | H | 6.28 |
| 19 | OH | H | $CH_3$ | H | 7.37 |
| 20 | OH | $OCH_3$ | $CH_3$ | H | 8.45 |
| 21 | $OCH_3$ | H | $CH_3$ | H | 6.53 |
| 22 | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2CH_3$ | 7.36 |
| 23 | $OCH_3$ | $OCH_3$ | $CH_3$ | $(CH_2)_2CH_3$ | 8.40 |
| 24 | $OCH_3$ | $OCH_3$ | $CH_3$ | $(CH_2)_3CH_3$ | 7.92 |
| 25 | $OCH_3$ | $OCH_3$ | $CH_3$ | $(CH_2)_4CH_3$ | 7.67 |
| 26 | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2CH=CH_2$ | 7.73 |
| 27[*] | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2CH=CHCH_3$ | 8.04 |
| 28 | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2CH=CHCH_2CH_3$ | 7.92 |
| 29 | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2C(CH_3)=CH_2$ | 7.96 |
| 30 | $OCH_3$ | $OCH_3$ | $CH_3$ | $CH_2CH=C(CH_3)_2$ | 8.16 |
| 31 | $OCH_3$ | $OCH_3$ | $CH_3$ |  | 7.31 |
| 32 | $OCH_3$ | $OCH_3$ | $CH_3$ |  | 7.97 |
| 33 | $OCH_3$ | $CH_3$ | $CH_3$ | H | 7.62 |
| 34[*] | $OCH_3$ | $CH=CH_2$ | $CH_3$ | H | 7.77 |
| 35 | $OCH_3$ | $CH_2CH_3$ | $CH_3$ | H | 8.03 |
| 36 | $OCH_3$ | $CH=CHCH_3$ | $CH_3$ | H | 8.30 |
| 37 | $OCH_3$ | $CH_2CH_2CH_3$ | $CH_3$ | H | 7.59 |
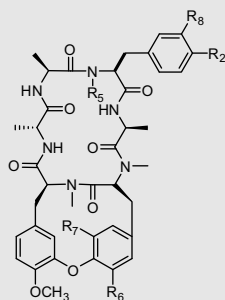
**Table 1** (continued)

| No. | $R_2$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $pIC_{50}$ |
|-----|-------|-------|-------|-------|-------|------------|
| 38 | $OCH_3$ | H | H | H | H | 6.92 |
| 39 | $OCH_3$ | $CH_3$ | OH | H | H | 5.32 |
| 40 | $OCH_3$ | $CH_3$ | H | OH | H | 5.24 |
| 41[*] | $OCH_3$ | $CH_3$ | H | H | OH | 6.67 |
| 42 | $OCH_3$ | $CH_3$ | OH | H | OH | 4.82 |
| 43 | OH | $CH_3$ | H | OH | H | 5.04 |
| 44 | $OCH_3$ | $CH_3$ | H | H | H | 8.77 |



| No. | $R_1$ | $R_3$ | X | Y | Z | W | $pIC_{50}$ |
|-----|-------|-------|---|---|---|---|------------|
| 45 | $OCH_3$ | $CH_3$ | S | O | O | O | 9.13 |
| 46 | $OCH_3$ | $CH_3$ | S | S | O | O | 8.67 |
| 47 | OH | $CH_3$ | S | O | O | O | 8.94 |
| 48 | $OCH_3$ | $CH_2COOCH_3$ | S | O | O | O | 8.32 |
| 49 | $OCH_3$ | $CH_3$ | O | S | O | O | 8.48 |
| 50[*] | $OCH_3$ | $CH_3$ | S | O | O | S | 8.26 |
| 51 | $OCH_3$ | $CH_3$ | S | O | S | O | 8.79 |
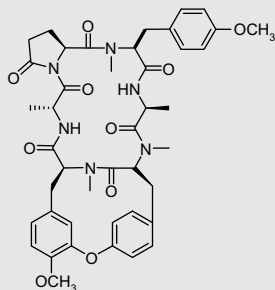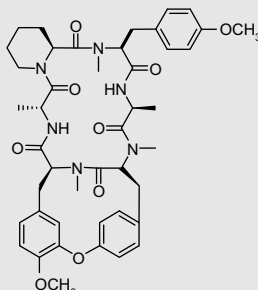


52 $pIC_{50}$=6.34    53 $pIC_{50}$=7.01    54 $pIC_{50}$=7.00
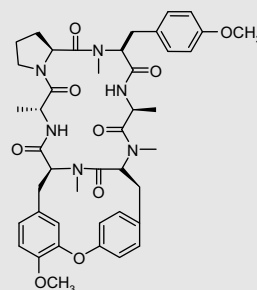
### 3.1. HQSAR analyses

21 HQSAR models were built based on different fragment distinction and fragment sizes using all of the 153 odd numbers of hologram lengths from 97 to 401 to make the model reliable.

Firstly, several combinations of fragment distinction parameters have been performed using the fragment size of 4–7 (Table 2) [21]. Chirality may take an important role in anti-tumor activity for RAs because the models showing a high cross-validated $q^2$ values included chirality. The best model 14 ($q^2 = 0.656$, $r^2 = 0.833$, SEE = 0.442) was obtained using Atoms/Bonds/Chirality as distinction information based on the hologram length of 59 and 5 components. Moreover, incorporating connections containing fragment into molecular hologram did not provide any improvement in these models as evaluated by $q^2$ and $r^2$ values.

Secondly, the influence of different fragment sizes over the models was further investigated based on the best model 14 and the results are summarized in Table 3 [22]. The best model 19 ($q^2 = 0.701$, $r^2 = 0.894$, SEE = 0.357) was obtained based on the fragment size of 6–9. The results demonstrated that bigger fragment size can provide an improvement in HQSAR model. This

result agreed with the special structure character of RAs which composed of six amino acid residues and linked with each other by resonance peptide bonds. Small fragment size will decrease the unity of amino acid residues. Moreover, the predictive power of this HQSAR model was validated by the test set molecules (Table 5). The graphic results for the experimental versus predicted $pIC_{50}$ values of both training and test sets are displayed in Fig. 6a.

Fig. 5 shows the most important fragments of compound **44** colored by the results of the final PLS analysis, which is one of the most potent inhibitors in the data set. Colors in red, red orange and orange reflect negative contributions, colors in yellow, green blue and green reflect positive contributions, and colors in white reflect intermediate contributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) Almost the whole cyclopeptide ring was colored by white, yellow and green blue, which indicated that positive contributions are mainly relative to its activity. Therefore this ring can serve as a structural scaffold for holding the pharmacophoric groups in suitable orientation. The alkyl hydrogen at $R_1$ is shown in yellow, showing positive contributions
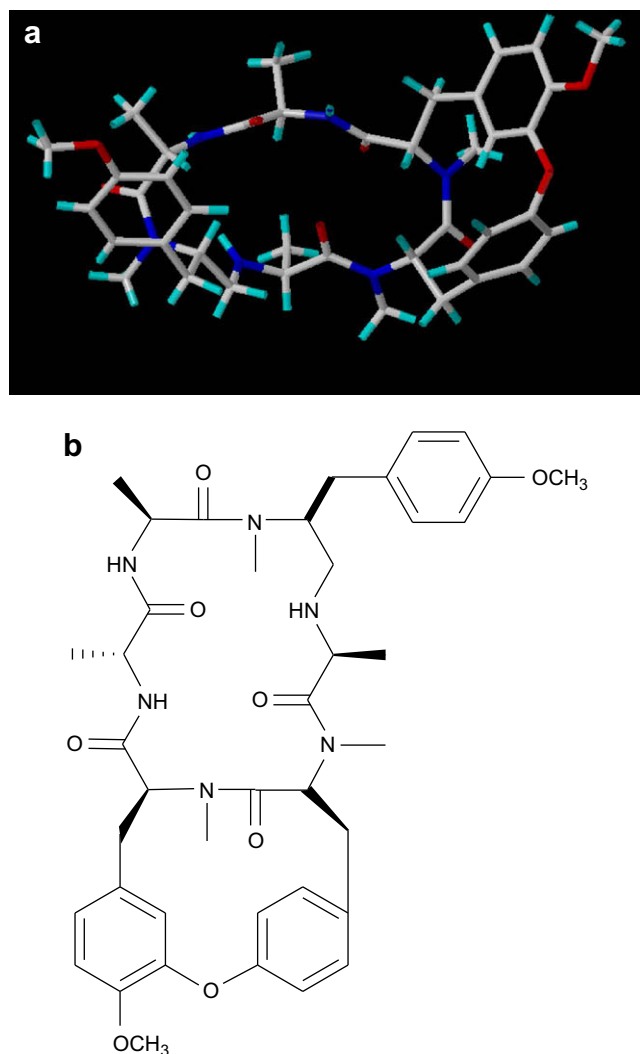
**Fig. 4.** 3D view of aligned 54 molecules (include training and test sets).

**Fig. 2.** Crystal structure of [Tyr-3–ψ(CH₂NH)–Ala-4]RA-VII.

**Table 2**

HQSAR analysis for various fragment distinction using fragment size of 4–7 and all the 153 odd numbers of hologram lengths from 97 to 401. A, Atoms; B, Bonds; C, Connections; D, Hydrogen Atoms; E, Chirality; F, Donor & Acceptor; $q^2$, LOO cross-validated correlation coefficient; $r^2$, non-cross-validated correlation coefficient; SEE, standard error of estimate; BL, best hologram length; N, number of components used in the PLS analysis.

| No. | Component | $q^2$ | $r^2$ | SEE | BL | N |
|---|---|---|---|---|---|---|
| 1 | A/B | 0.633 | 0.802 | 0.481 | 85 | 5 |
| 2 | A/B/C | 0.627 | 0.753 | 0.524 | 75 | 3 |
| 3 | A/B/C/D | 0.562 | 0.814 | 0.466 | 241 | 5 |
| 4 | A/B/C/E | 0.606 | 0.713 | 0.559 | 313 | 2 |
| 5 | A/B/C/D/E | 0.599 | 0.845 | 0.430 | 185 | 6 |
| 6 | A/C/F | 0.520 | 0.633 | 0.639 | 73 | 3 |
| 7 | A/B/C/D/F | 0.468 | 0.655 | 0.620 | 163 | 3 |
| 8 | A/B/D | 0.619 | 0.830 | 0.451 | 199 | 6 |
| 9 | A/B/D/F | 0.491 | 0.703 | 0.582 | 173 | 4 |
| 10 | A/B/C/F | 0.504 | 0.635 | 0.638 | 119 | 3 |
| 11 | A/B/E/F | 0.575 | 0.786 | 0.497 | 221 | 4 |
| 12 | A/B/D/E | 0.628 | 0.858 | 0.412 | 301 | 6 |
| 13 | A/B/F | 0.521 | 0.631 | 0.641 | 111 | 3 |
| 14 | A/B/E | 0.656 | 0.833 | 0.442 | 59 | 5 |

to the biological activity. This is supported by the fact that when the OMe group in this position is replaced by OH, there is a decrease in its potency, i.e., comparison of **20** (pIC₅₀ = 8.45) with **44** (pIC₅₀ = 8.77).
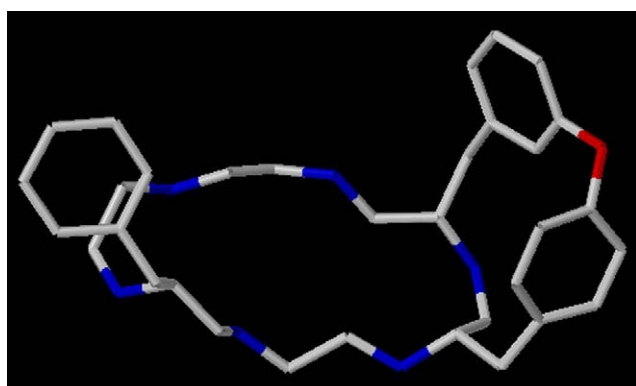
**Table 3**

HQSAR analysis for various fragment size using the best fragment distinction (Atoms/Bonds/Chirality) and all the 153 odd numbers of hologram lengths from 97 to 401. $q^2$, LOO cross-validated correlation coefficient; $r^2$, non-cross-validated correlation coefficient; SEE, standard error of estimate; BL, best hologram length; N, number of components used in the PLS analysis.

| No. | Fragment size | $q^2$ | $r^2$ | SEE | BL | N |
|---|---|---|---|---|---|---|
| 15 | 2–5 | 0.598 | 0.666 | 0.603 | 53 | 2 |
| 16 | 3–6 | 0.575 | 0.685 | 0.586 | 123 | 2 |
| 17 | 4–7 | 0.656 | 0.833 | 0.442 | 59 | 6 |
| 18 | 5–8 | 0.639 | 0.738 | 0.540 | 97 | 3 |
| 19 | 6–9 | 0.701 | 0.894 | 0.357 | 93 | 6 |
| 20 | 7–10 | 0.707 | 0.843 | 0.428 | 95 | 5 |
| 21 | 8–11 | 0.699 | 0.870 | 0.390 | 95 | 5 |

### 3.2. CoMFA analyses

The steric and electrostatic CoMFA fields yielded a cross-validated $q^2 = 0.510$ with 8 components, non-cross-validated $r^2$ of 0.964, SEE = 0.213 and F value of 127.358. The contribution of steric and electrostatic fields is 51.1% and 48.9%, respectively. The graph of experimental versus predicted pIC₅₀ values of the training and test set compounds is depicted in Fig. 6b. Steric and
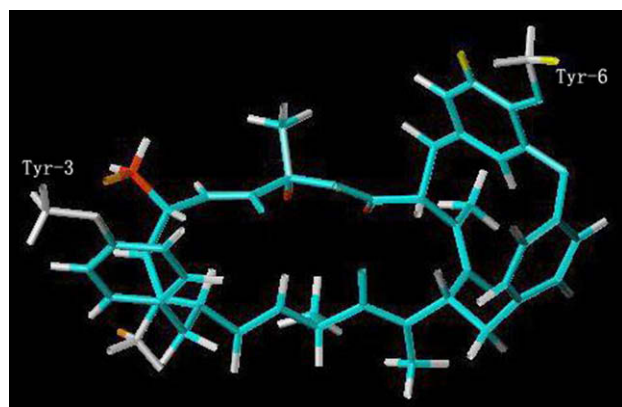


**Fig. 3.** Common fragment used in database alignment.

**Fig. 5.** HQSAR contribution map of compound **44**.

electrostatic field contour maps are shown in Fig. 7 with compound **44**.

In CoMFA steric map (Fig. 7a), green contours indicate regions where steric bulk groups increase the activity, while yellow
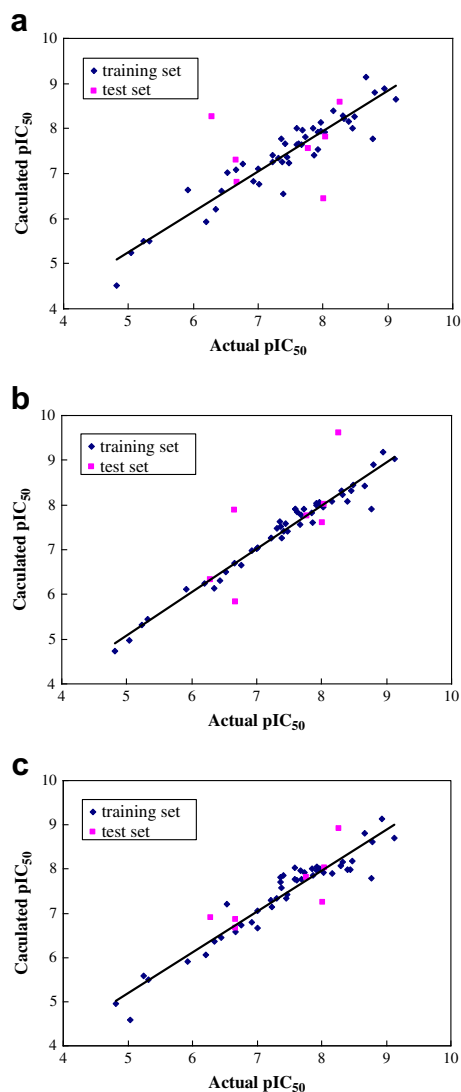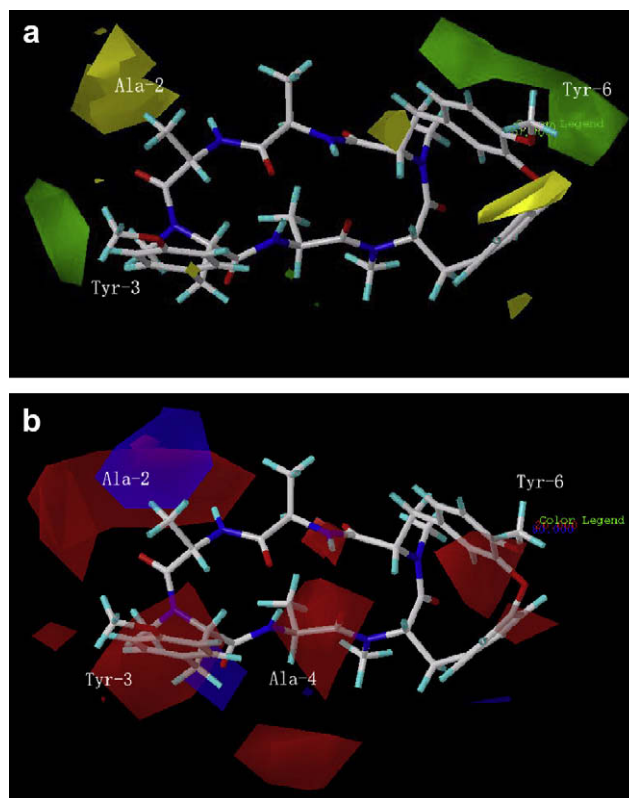


**Fig. 7.** CoMFA STDEV*COEFF contour maps. (a) Steric fields: green contours indicate regions where bulky groups increase activity, while yellow contours indicate regions where bulky groups decrease activity; and (b) electrostatic fields: blue contours indicate regions where electron positive groups increase activity, while red contours indicate regions where electron negative groups increase activity. Compound **44** is displayed as a reference.



**Fig. 6.** Graph of experimental versus calculated pIC$_{50}$ values from HQSAR (a), CoMFA (b) and CoMSIA (c) analyses for the training and test set compounds.

contours indicate regions where steric bulk groups decrease the activity. There is one green contour around R$_2$ position, which can be explained by compound **44** (pIC$_{50}$ = 8.77) in which R$_2$ position is substituted by methoxyl and compound **21** (pIC$_{50}$ = 6.53) by hydrogen. There is another green region located at the R$_1$. This also explains why compound **20** (pIC$_{50}$ = 8.45) is less active than compound **44** (pIC$_{50}$ = 8.77) when hydroxyl is substituted by methoxyl. The yellow contour near the R$_3$ position depicted that the steric occupancy with more bulky groups in this region will decrease in activity. For **4** (pIC$_{50}$ = 7.85), **15** (pIC$_{50}$ = 7.22), **8**

**Table 4**
Summary of HQSAR, CoMFA, and CoMSIA models' results. q$^2$, LOO cross-validated correlation coefficient; r$^2$, non-cross-validated correlation coefficient; N, number of components used in the PLS analysis; SEE, standard error of estimate; F value, F-statistic for the analysis; BL, best hologram length.

| Component | HQSAR | CoMFA | CoMSIA |
|---|---|---|---|
| q$^2$ | 0.701 | 0.510 | 0.613 |
| r$^2$ | 0.894 | 0.964 | 0.920 |
| N | 6 | 8 | 6 |
| BL | 93 | – | – |
| F value | – | 127.358 | 72.927 |
| SEE | 0.357 | 0.213 | 0.309 |
| Steric | – | 0.511 | 0.136 |
| Electrostatic | – | 0.489 | – |
| Hydrophilic | – | – | 0.180 |
| Donor | – | – | 0.413 |
| Acceptor | – | – | 0.271 |

**Table 5**
Experimental and predicted $pIC_{50}$ values of training and test set (labeled with an asterisk) compounds.

| No. | Experimental | HQSAR Predicted | CoMFA Predicted | CoMSIA Predicted |
|---|---|---|---|---|
| 1[*] | 6.66 | 7.31 | 7.89 | 6.85 |
| 2 | 7.23 | 7.41 | 7.26 | 7.15 |
| 3 | 7.38 | 6.56 | 7.26 | 7.57 |
| 4 | 7.85 | 8.00 | 7.82 | 8.02 |
| 5[*] | 8.01 | 6.44 | 7.61 | 7.25 |
| 6 | 6.20 | 5.93 | 6.25 | 6.06 |
| 7 | 6.44 | 6.62 | 6.30 | 6.46 |
| 8 | 6.76 | 7.22 | 6.66 | 6.73 |
| 9 | 7.41 | 7.66 | 7.40 | 7.85 |
| 10 | 7.47 | 7.23 | 7.42 | 7.42 |
| 11 | 6.66 | 7.09 | 6.71 | 6.58 |
| 12 | 7.86 | 7.41 | 7.61 | 7.86 |
| 13 | 7.45 | 7.36 | 7.59 | 7.34 |
| 14 | 5.92 | 6.64 | 6.12 | 5.92 |
| 15 | 7.22 | 7.25 | 7.25 | 7.29 |
| 16 | 7.59 | 7.65 | 7.90 | 7.77 |
| 17 | 7.69 | 7.95 | 7.78 | 7.76 |
| 18[*] | 6.28 | 8.26 | 6.34 | 6.91 |
| 19 | 7.37 | 7.27 | 7.52 | 7.70 |
| 20 | 8.45 | 8.01 | 8.31 | 7.99 |
| 21 | 6.53 | 7.03 | 6.50 | 7.21 |
| 22 | 7.36 | 7.78 | 7.62 | 7.82 |
| 23 | 8.40 | 8.15 | 8.09 | 7.99 |
| 24 | 7.92 | 7.91 | 8.04 | 8.06 |
| 25 | 7.67 | 7.64 | 7.56 | 7.96 |
| 26 | 7.73 | 7.81 | 7.90 | 7.91 |
| 27[*] | 8.04 | 7.82 | 8.02 | 8.02 |
| 28 | 7.92 | 7.53 | 7.98 | 7.97 |
| 29 | 7.96 | 7.95 | 8.03 | 7.99 |
| 30 | 8.16 | 8.39 | 8.07 | 7.89 |
| 31 | 7.31 | 7.34 | 7.48 | 7.34 |
| 32 | 7.97 | 8.15 | 8.06 | 8.01 |
| 33 | 7.62 | 7.67 | 7.85 | 7.74 |
| 34[*] | 7.77 | 7.56 | 7.75 | 7.82 |
| 35 | 8.03 | 7.93 | 7.95 | 7.92 |
| 36 | 8.30 | 8.29 | 8.33 | 8.06 |
| 37 | 7.59 | 8.00 | 7.90 | 8.02 |
| 38 | 6.92 | 6.83 | 6.97 | 6.80 |
| 39 | 5.32 | 5.49 | 5.45 | 5.49 |
| 40 | 5.24 | 5.49 | 5.32 | 5.59 |
| 41[*] | 6.67 | 6.80 | 5.84 | 6.66 |
| 42 | 4.82 | 4.52 | 4.74 | 4.95 |
| 43 | 5.04 | 5.24 | 4.98 | 4.59 |
| 44 | 8.77 | 7.77 | 7.91 | 7.79 |
| 45 | 9.13 | 8.64 | 9.02 | 8.71 |
| 46 | 8.67 | 9.13 | 8.42 | 8.82 |
| 47 | 8.94 | 8.88 | 9.17 | 9.13 |
| 48 | 8.32 | 8.23 | 8.23 | 8.17 |
| 49 | 8.48 | 8.26 | 8.45 | 8.19 |
| 50[*] | 8.26 | 8.58 | 9.62 | 8.92 |
| 51 | 8.79 | 8.79 | 8.90 | 8.62 |
| 52 | 6.34 | 6.21 | 6.13 | 6.37 |
| 53 | 7.01 | 6.77 | 7.04 | 7.05 |
| 54 | 7.00 | 7.10 | 7.02 | 6.68 |

**Table 6**
Summary of CoMSIA results. A, Steric field; B, Electrostatic field; C, Hydrophobic field; D, Donor field; E, Acceptor field; $q^2$, LOO cross-validated correlation coefficient; $r^2$, non-cross-validated correlation coefficient; N, number of components used in the PLS analysis; SEE: standard error of estimate; F value, F-statistic for the analysis.

| No. | CoMSIA field | $q^2$ | N | $r^2$ | SEE | F value |
|---|---|---|---|---|---|---|
| 1 | A/B/C/D/E | 0.568 | 5 | 0.885 | 0.367 | 63.213 |
| 2 | A/B/C/D | 0.569 | 6 | 0.915 | 0.320 | 71.504 |
| 3 | A/B/C/E | 0.527 | 6 | 0.907 | 0.334 | 64.967 |
| 4 | A/C/D/E | 0.613 | 6 | 0.920 | 0.309 | 72.927 |
| 5 | A/B/D/E | 0.552 | 5 | 0.882 | 0.371 | 61.496 |
| 6 | B/C/D/E | 0.552 | 5 | 0.883 | 0.270 | 61.927 |
| 7 | A/B/C | 0.526 | 6 | 0.894 | 0.356 | 56.401 |
| 8 | A/B/D | 0.469 | 6 | 0.899 | 0.347 | 59.621 |
| 9 | B/C/D | 0.523 | 5 | 0.878 | 0.379 | 58.740 |
| 10 | B/C/E | 0.508 | 6 | 0.907 | 0.334 | 65.174 |
| 11 | A/C/D | 0.624 | 5 | 0.896 | 0.349 | 70.620 |
| 12 | A/C/E | 0.540 | 6 | 0.914 | 0.321 | 70.838 |

### 3.3. CoMSIA analyses

12 CoMSIA models were generated using different combined fields. The statistical parameters are summarized in Tables 4 and 6. Models including hydrogen-bond donor and acceptor fields showed a high cross-validated $q^2$ value, which indicated that hydrogen bonds are important factors for RAs' anti-tumor activity. Model 4 with the combinations of steric, hydrophobic, hydrogen-bond donor and acceptor fields yielded the highest $q^2$ (0.613) with 6 components, and $r^2$ (0.920) with SEE of 0.309. The contributions of steric, hydrophobic, hydrogen-bond donor and acceptor fields are 13.6%, 18.0%, 41.3%, and 27.1%, respectively. Fig. 6c shows the experimental versus predicted $pIC_{50}$ values in the training and test set compounds. CoMSIA steric, hydrophobic, hydrogen-bond donor and acceptor field contour maps are shown in Fig. 8 with compound **44**. Since the steric contour of CoMSIA (Fig. 8a) is very similar with that of CoMFA, only hydrophobic and hydrogen-bond fields will be described as follows.

In hydrophobic map (Fig. 8b), yellow contours suggest that a hydrophobic substituent may favor anti-tumor activity, while white contours unfavor its activity. A big yellow contour near the $R_3$ suggests that hydrophobic substitutes would favor the anti-tumor activity. For compounds **10** ($pIC_{50} = 7.47$), **15** ($pIC_{50} = 7.22$) and **14** ($pIC_{50} = 5.92$) where $R_3$ was substituted by aldehyde, hydroxyl and carboxylic groups, respectively, result shows the same order between anti-tumor activity and hydrophobic substitutes. Moreover, yellow contours near the phenyl portion of Tyr-6 and Tyr-3 can be explained by analyzing the structural features and biological activities of compounds **44** and **40**, **41**. Replacement of hydrogen (**44**) by hydroxyl (**40**, **41**) at the $R_7$ and $R_8$ positions decreased the activity from 8.77 to 5.24 and 6.67 of $pIC_{50}$ values.

H-bond donor and acceptor contour maps (Fig. 8c and d) demonstrated that H-bond donors (amino hydrogen) and H-bond acceptors (carboxyl oxygen) of RAs are very important particularly because RAs possess a stable antiparallel β-sheet conformation formed by two intra-molecular hydrogen bonds between Ala-1 and Ala-4. These intra-molecular hydrogen bonds make the cyclic ring more rigid and play an important role in maintaining pharmacophore conformation. In Fig. 8c, there is a big cyan contour covering the amino hydrogen of Ala-1 and Ala-4 which indicates that hydrogen-bond donors (amino hydrogen) in these regions favor the anti-tumor activity. In Fig. 8d, white contours located at the position of Z and the carboxyl oxygen of Ala-4 suggest that the presence of hydrogen-bond acceptor (carboxyl
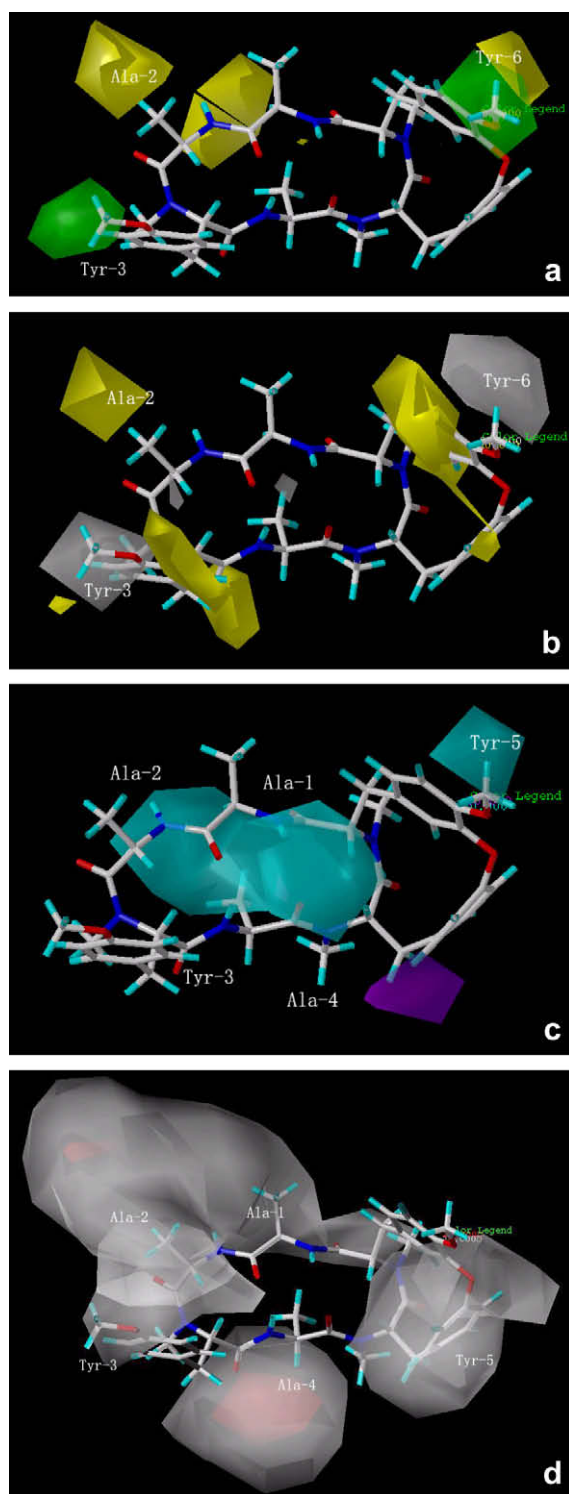
($pIC_{50} = 6.76$) and **11** ($pIC_{50} = 6.66$), the $pIC_{50}$ values decreased according to the order of $\mathbf{4} > \mathbf{15} > \mathbf{8} > \mathbf{11}$ with increasing $CH_2$ number from 1 to 4.

For CoMFA electrostatic map (Fig. 7b), blue contours indicate regions where electron positive groups increase activity, and red contours indicate regions where electron negative groups increase activity. There is one blue contour around the $R_3$ position, which can be explained by compounds **1** ($pIC_{50} = 6.66$) and **2** ($pIC_{50} = 7.23$), which possess carboxyl groups at the $R_3$ position with reduced activity. A red contour located at $R_2$ position, which can be explained by compound **44** ($pIC_{50} = 8.77$) and compound **21** ($pIC_{50} = 6.53$). Compound **44** possesses one more electron negative group methoxyl at this position, but compound **21** has hydrogen here.

**Fig. 8.** CoMSIA STDEV*COEFF contour maps. (a) Steric fields: green contours and yellow contours show regions where an increase in bulky groups will increase and decrease activity, respectively; (b) hydrophobic fields: yellow contours and white contours show regions where an increase in lipophilicity and hydrophilicity will increase activity, respectively; (c) H-bond donor contour map: cyan contour indicates regions where hydrogen-bond donor groups increase activity; (d) H-bond acceptor contour map: white contour indicates regions where hydrogen-bond acceptor groups increase activity. Compound **44** is displayed as a reference.

oxygen) at these two positions favors the anti-tumor activity. An instance is that the anti-tumor activity of compound **44** will be lost after the carboxyl oxygen is substituted by two hydrogens at the X position [18].

## 4. Conclusion

In this work, HQSAR, CoMFA and CoMSIA methods were employed to develop 2D- and 3D-QSAR models for 54 anti-tumor RAs. The good correlation between experimental and predicted $pIC_{50}$ values for the test set compounds proved the reliability of these QSAR models. HQSAR ($q^2 = 0.701$, $r^2 = 0.894$) result showed that chirality descriptor plays an important role in anti-tumor activity of RAs and OMe at $R_1$ and $R_2$ is necessary for increasing their activity. CoMFA ($q^2 = 0.510$, $r^2 = 0.964$) and CoMSIA ($q^2 = 0.613$, $r^2 = 0.920$) results demonstrated that small bulky and electropositive side chains at $R_3$ position and hydrophobic groups at $R_7$ and $R_8$ positions will increase their activity. Moreover, intra-molecular hydrogen bonds between residues 1 and 4 are necessary to maintain the pharmacophoric conformation of RAs. These results may be helpful in designing novel and potential anti-tumor RAs.

## References

[1] N.H. Tan, J. Zhou, Chem. Rev. 106 (2006) 840–895.
[2] H. Fujiwara, S. Saito, Y. Hitotsuyanagi, K. Takeya, Y. Ohizumi, Cancer Lett. 209 (2004) 223–229.
[3] H.L. Li, Z.T. Gao, L. Kang, H.L. Zhang, K. Yang, K.Q. Yu, X.M. Luo, W.L. Zhu, K.X. Chen, J.H. Shen, X.C. Wang, H.L. Jiang, Nucleic Acids Res. 34 (2006) 219–224.
[4] SYBYL, Version 6.9, Tripos Inc., St. Louis, MO.
[5] R.D. Cramer III, D.E. Patterson, J.D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[6] M. Clark, R.D. Cramer III, N.V. Opdenbosch, J. Comput. Chem. 10 (1989) 982–1012.
[7] T.L. Moda, C.A. Montanarib, A.D. Andricopulo, Bioorg. Med. Chem. 15 (2007) 7738–7745.
[8] H. Itokawa, T. Yamamiya, H. Morita, K. Takeya, J. Chem. Soc., Perkin Trans. 1 (1992) 455–459.
[9] H. Itokawa, K. Saitou, H. Morita, K. Takeya, K. Yamada, Chem. Pharm. Bull. 40 (1992) 2984–2989.
[10] H. Itokawa, K. Kondo, Y. Hitotsuyanagi, K. Takeya, Heterocycles 36 (1993) 1837–1843.
[11] H. Itokawa, K. Kondo, Y. Hitotsuyanagi, M. Isomura, K. Takeya, Chem. Pharm. Bull. 41 (1993) 1402–1410.
[12] H. Itokawa, K. Kondo, Y. Hitotsuyanagi, A. Nakamura, H. Morita, K. Takeya, Chem. Pharm. Bull. 41 (1993) 1266–1269.
[13] Y. Hitotsuyanagi, J. Suzuki, Y. Matsumoto, K. Takeya, H. Itokawa, J. Chem. Soc., Perkin Trans. 1 (1994) 1887–1889.
[14] Y. Hitotsuyanagi, J. Suzuki, K. Takeya, H. Itokawa, Bioorg. Med. Chem. Lett. 4 (1994) 1633–1636.
[15] Y. Hitotsuyanagi, S. Lee, I. Ito, K. Kondo, K. Takeya, T. Yamagishi, T. Nagate, H. Itokawa, J. Chem. Soc., Perkin Trans. 1 (1996) 213–217.
[16] Y. Hitotsuyanagi, Y. Matsumoto, S. Sasaki, J. Suzuki, K. Takeya, K. Yamaguchi, H. Itokawa, J. Chem. Soc., Perkin Trans. 1 (1996) 1749–1755.
[17] H.Q. Huang, X.L. Pan, N.H. Tan, G.Z. Zeng, C.J. Ji, Eur. J. Med. Chem. 42 (2007) 365–372.
[18] J.J.P. Stewart, J. Comput. Chem. 10 (2) (1989) 209–220.
[19] http://www.sparkle.pro.br/.
[20] W. Tong, D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sheehan, J. Chem. Inf. Comput. Sci. 38 (1998) 669–677.
[21] G. Klebe, U. Abraham, T. Mietener, J. Am. Chem. Soc. 37 (1994) 4130–4136.
[22] C.R. Rodrigues, T.M. Flaherty, C. Springer, J.H. McKerrow, F.E. Cohen, Bioorg. Med. Chem. Lett. 12 (2002) 1537–1541.