Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 2159–2170
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

QSAR, action mechanism and molecular design of flavone and isoflavone derivatives with cytotoxicity against HeLa

Si Yan Liao, Jin Can Chen, Li Qian, Yong Shen, Kang Cheng Zheng*

School of Chemistry and Chemical Engineering, Zhongshan (Sun Yat-Sen) University, Guangzhou 510275, People's Republic of China

Received 22 June 2007; received in revised form 26 October 2007; accepted 29 October 2007
Available online 7 November 2007

Abstract

The quantitative structure–activity relationship (QSAR) of 32 flavone and isoflavone derivatives with cytotoxicity expressed as pGC_{50}, which is defined as the negative value of the logarithm of necessary molar concentration of this series of compounds to cause 50% growth inhibition against the human cervical epithelioid carcinoma cell line (HeLa), has been studied by using the density functional theory (DFT), molecular mechanics (MM2) and statistical methods. In order to obtain QSAR model with high predictive ability, the original dataset was randomly divided into a training set comprising 26 compounds and a test set comprising the rest 6 compounds. An optimal model for the training set with significant statistical quality ($R^2_A = 0.852$) and predictive ability ($q^2 = 0.818$) was established. The same model was further applied to predict pGC_{50} values of the 6 compounds in the test set, and the resulting predictive correlation coefficient R^2_{pred} reaches 0.738, further showing that this QSAR model has high predictive ability. It is very interesting to find that the cytotoxicities of these compounds against HeLa appear to be mainly governed by two quantum-chemical factors, i.e., the energy (E_{LUMO}) of the lowest unoccupied molecular orbital (LUMO) and the net charges of C atom at site 6 on aromatic rings (Q_{C6}). Here the possible action mechanism of these compounds was analyzed and discussed in detail, in particular, the fact why the flavone derivatives have considerably higher cytotoxicity than isoflavone derivatives was reasonably explained. Based on this QSAR equation, 5 new compounds with higher cytotoxicity have been theoretically designed. Such results can offer useful theoretical references for experimental works.

Keywords: QSAR; Molecular design; Flavone derivative; Cytotoxicity; DFT calculation

1. Introduction

Flavonoids and isoflavonoids are a class of natural product that shows extensive biological activities with low toxicity, such as antiviral, antiinflammatory, antioxidant and antitumoral actions [1–3]. These compounds are distributed in a wide variety of plants and commonly consumed with the human diet. Hence, their use as potential therapeutic compounds against a broad range of diseases is of interest [4]. Many of the pharmacological properties of these compounds have been found to relate to their abilities to inhibit enzymes involved in cell activation, such as cAMP phosphodiesterases, kinases, topoisomerases and other regulatory enzymes [5].

One of the most interesting biological properties of some flavonoids is their cytotoxicity. Several polyalkoxylated flavonoid aglycones have shown interesting antitumor properties, including flavone [6], flavonol [7], and flavanone [8]. Chen et al. proposed that the cytotoxicity induced by flavone is mediated by inducing the occurrence of apoptosis characterized by the appearance of DNA ladders, apoptotic bodies and hypodiploid cells [9]. The DNA-binding affinity of the drug is one of significant factors that may affect the cytotoxicity. Recently, Wang et al. have reported preparation of certain oxime- and methyloxime-containing flavone and isoflavone derivatives and investigated their antiproliferative activities [10]. They expected the corresponding oximes (H-bond donors) and *o*-methyloximes (H-bond acceptors) to form H-bonds with DNA upon intercalation [11]. Evaluation of the antiproliferative activities of these compounds showed a potent inhibitory

* Corresponding author. Tel.: +86 20 84110696; fax: +86 20 84002245.
E-mail address: ceszkc@mail.sysu.edu.cn (K.C. Zheng).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.10.033

effect on the cell growth in all 60 human tumor cell lines [10]. These discoveries exemplified the great potential of developing flavone and isoflavone derivatives as a new class of anticancer drug. However, so far the action mechanisms between these compounds and DNA remain unclarified at the molecular level. Fortunately, the availability of experimental cytotoxicity data on flavone and isoflavone derivatives affords us an opportunity to apply quantitative structure–activity relationship (QSAR) approach for obtaining some significant insights into the action mechanism and molecular design of these compounds.

For many years, QSARs have been efficiently used for the study of biological mechanisms of various reactive chemicals [12,13]. This is a powerful technique, which quantitatively relates variations in biological activity to the molecular structures or properties. In order to establish an excellent QSAR equation, it is very important to select high-quality descriptors, because the success of a QSAR model is highly dependent on the choice of effective descriptors. Recently, a number of quantum-chemical descriptors like charges, orbital energies, frontier orbital densities, and dipole moment, etc. obtained from density functional theory (DFT) calculations have been successfully used in developing different QSARs for predicting reactivity in terms of the structures and physicochemical properties of molecules [14]. In addition, molecular mechanics (MM2) calculations providing molecular properties descriptors like steric and hydrophobic coefficients have also been used in QSAR study [15].

In this paper, a QSAR of a series of flavone and isoflavone derivatives with anticancer activity against human cervical epithelioid carcinoma cell line (HeLa) in literature [10] was studied by using a combined method of the DFT, MM2, and statistics, and a robust and predictive QSAR model only with two main independent quantum-chemical descriptors was established. In particular, based on this QSAR model, the possible mechanism regarding these compounds interacting with DNA was explored, and some new compounds with higher cytotoxicity against HeLa were theoretically designed.

2. Calculations

2.1. Computational methods

The structural schematic diagrams of compounds used in model building are shown in Fig. 1 and Table 1. The molecular mechanics MM2 method in Chem3D software [16] was applied to search for lower energy conformations for each molecule. Then the stable geometry structures of these compounds were further fully optimized using the DFT calculations at the level of Becke's three-parameter hybrid functional (B3LYP) [17] and 6-31G basis set. Via the frequency calculations, there was not an imaginary frequency appearance for all configurations at the energy minima showing that the optimized stable structures are reasonable and reliable. Atomic charges were computed using the natural population analysis (NPA) [18]. The all quantum-chemical calculations were performed by Gaussian 03 program package [19]. In addition, the Hyperchem software [20] was also employed to calculate the following parameters of molecular properties from the energy-minimized structures: surface area (S), volume (V), hydrophobic coefficient ($\log P$), molar refractivity (MR), and so on.

2.2. Statistical analysis

From the results of the DFT calculations, the quantum-chemical descriptors were obtained for the model building as follows: the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}), the energy difference between the LUMO and the HOMO (ΔE_{L-H}), the total dipole moment (μ) of the molecule, the net charges of C atom at site 6 on aromatic rings (Q_{C6}), the total net charges of R_1, R_2, and R_3 (Q_{R1}, Q_{R2}, Q_{R3},), etc. Meanwhile, the descriptors of molecular properties, such as the hydrophobic coefficient and molar refractivity of substituents R_1, R_2, and R_3 ($\log P_{R1}$, $\log P_{R2}$, $\log P_{R3}$; MR_{R1}, MR_{R2}, MR_{R3}; respectively), were also selected. More than 20 descriptors were adopted as candidates for the correlation analysis.

The cytotoxicities of 32 studied compounds are expressed in terms of GC_{50}, which is necessary molar concentration of compound causing 50% cell growth inhibition against the human cervical epithelioid carcinoma cell line (HeLa); the corresponding values are presented in Table 1. All original GC_{50} values are usually converted to negative logarithm of GC_{50} (pGC_{50}) in QSAR study.

In order to select out the predominant descriptors affecting the cytotoxic activity, the correlation analysis was performed by the statistical software SPSS [21]. To eliminate the intercorrelative parameters and minimize the information overlap in the model, the descriptors with lower inter-correlation ($|r| < 0.5$) were only considered [22]. The descriptors with higher correlation to the pGC_{50} and lower inter-correlation were selected to carry out the stepwise multiple linear regression analysis to establish the optimal QSAR equation.

In the next step, the predictive ability of the equation was evaluated by a well-known method of "leave-one-out" (LOO) cross-validation [23]. The square of cross-validation coefficient q^2, which is used as a criterion of both robustness and predictive ability of the model, should be greater than 0.5 for a reliable model [24].

It is commonly accepted that the internal validation of QSAR model built from training set is sufficient to confirm its predictive power. However, a high q^2 in training set only shows a good internal validation, but it does not automatically infer its high predictive ability for an external test set [25], because usually q^2 overestimates the predictive ability of obtained model. In order to obtain QSAR with more reliable predictive ability, external validation is also crucial. Thus, the dataset of 32 compounds was divided into a training set comprising randomly selected 26 compounds and a test set comprising the rest 6 compounds. QSAR model was built from training set, and it was determined with the test set to confirm its predictive ability. A QSAR model is considered

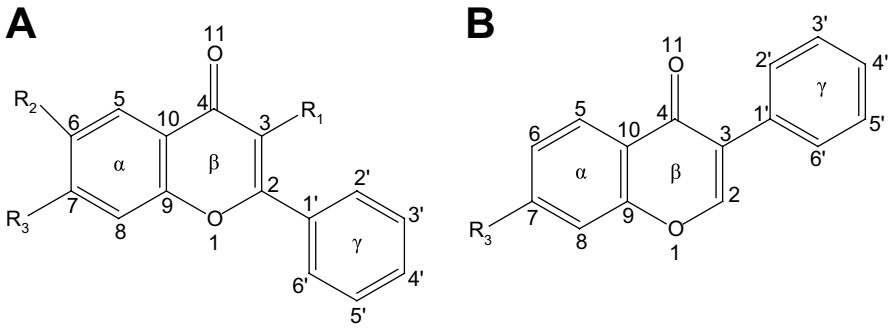


Fig. 1. Molecular structures and numbering of flavone (A) and isoflavone (B) derivatives in the present study.

to own high predictive power only if its square of predictive correlation coefficient (R^2_{pred}) between the experimental and predicted activities is greater than 0.6 for the test set [26].

3. Results and discussion

3.1. QSAR equation

Considering the balance of the QSAR quality and the number of employed descriptors, an optimal QSAR equation was obtained for 26 compounds in training set through multiple linear regression analysis as follows:

$$pGC_{50} = 2.637 - 42.90E_{LUMO} + 0.371Q_{C6} \quad (1)$$

$n = 26$; $R^2 = 0.828$; $R^2_A = 0.813$; $r = 0.126$; $S_{reg} = 0.152$; $S_{pre} = 0.143$; $q^2 = 0.780$; $F = 55.21$; $p < 0.0001$

where n is the number of compounds in training set; R^2 is the square of correlation coefficient of regression; R^2_A is the square of adjusted correlation coefficient; r is the root-mean-square deviation; S_{reg} is the standard deviation of regression and S_{pre} is that of the prediction; F is the Fisher's F value using the F statistics; p is the p value using the F statistics; and q^2 is the square of LOO cross-validation coefficient. A good QSAR model has characters of large F, small r and S, low p value, R^2 and q^2 values close to 1, as well as $p < 0.001$. So the above established QSAR model shows significant statistical quality. The selected parameters in Eq. (1) and the deviations of regression and prediction are listed in Table 2.

From Table 2, we can see that, compound **27** has the largest regression deviation (0.32) and the largest prediction deviation (0.34), thus compound **27** may be regarded as outlier. We have proposed a new approach to determine the outliers using "LOO" cross-validation coefficient q^2_{n-i}, which denotes the square of LOO cross-validation coefficient of compound i computed by the new cross-validation procedure after leaving this datum point (no. i) out from n compounds. The compound with unduly high q^2_{n-i} value can be considered as an outlier, and the compound with the low value can be indicated an influential point [27]. The q^2_{n-i} values of this series of compounds are also listed in Table 2. From Table 2, it is clearly found that compound **27** has too large q^2_{n-i} value (0.818) in the 26 compounds as training set, so compound **27** can be confirmed to be the outlier. In addition, via a careful analysis of the correlation between the pGC_{50} values and the molecular structures for this series of compounds, we find an interesting fact as follows: all flavone derivatives except for compound **27** have higher cytotoxicity than any one of isoflavone derivatives (the reasons will be explained in Section 3.2). It further makes sure that compound **27** is the outlier. After omitting compound **27**, an optimal model Eq. (2) is obtained as follows:

$$pGC_{50} = 2.618 - 43.30E_{LUMO} + 0.336Q_{C6} \quad (2)$$

$n = 25$; $R^2 = 0.865$; $R^2_A = 0.852$; $r = 0.110$; $S_{reg} = 0.136$; $S_{pre} = 0.128$; $q^2 = 0.818$; $F = 70.35$; $p < 0.0001$

Comparing Eq. (2) with Eq. (1), we find that Eq. (2) with larger R^2, q^2 and F values as well as smaller r and S values indicates the best statistical quality, and that using q^2_{n-i} to determine the outlier is more feasible. So we will only discuss the obtained optimal model Eq. (2) below.

The deviations of regression and prediction of Eq. (2) are also listed in Table 2, and the correlation matrix between pGC_{50} and selected parameters in Eq. (2) is listed in Table 3, which shows that the selected parameters are independent with small inter-correlation coefficients ($|r| < 0.35$). The results show that the established QSAR model Eq. (2) is of robustness.

A well-accepted QSAR model should be able to accurately predict activities of new compounds. Therefore, the QSAR model Eq. (2) was further applied to predict the cytotoxicities of 6 compounds in the test set. From Table 4, we can find that the predicted deviations lie in a range of −0.19 to 0.28, which is very near to the range (−0.26 to 0.21) of regression deviations for Eq. (2) in Table 2, and the resulting predictive correlation coefficient R^2_{pred} reaches 0.738. Hence, the results obtained from the test set further demonstrate that this QSAR model Eq. (2) is very robust and predictive, and can offer some useful theoretical references for understanding the action mechanism and directing the molecular design of this kind of compound with high cytotoxicity. The plot of the predicted pGC_{50} values based on Eq. (2) versus experimental ones is shown in Fig. 2. Obviously, the predicted pGC_{50} values are in a good agreement with experimental ones.

Table 1
Structures and experimental cytotoxicities (against human cervical epithelioid carcinoma cell line) of the flavone and isoflavone derivatives (Ref. [10])

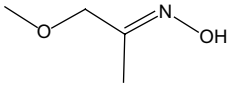
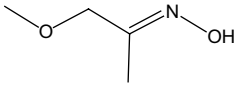
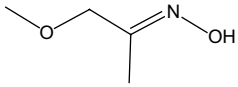
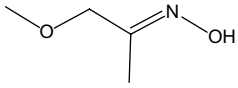
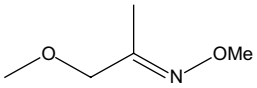
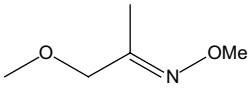
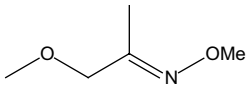
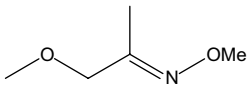
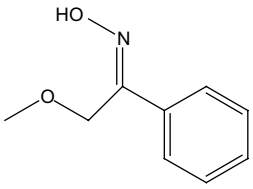
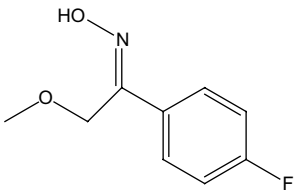
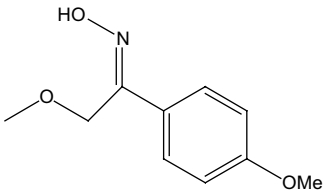
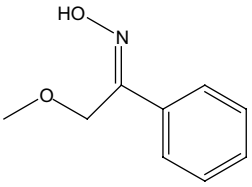
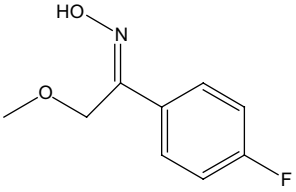
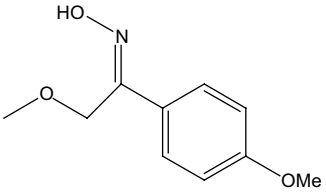
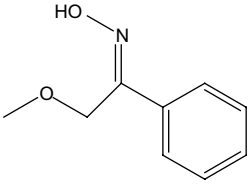
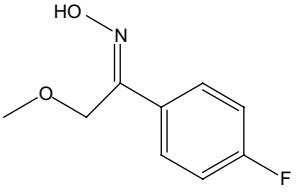
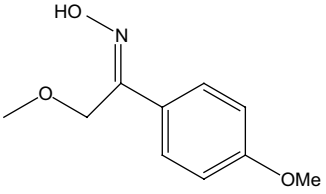
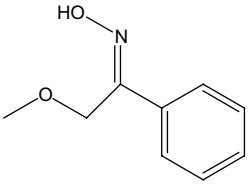
No.	Form	R_1	R_2	R_3	GC_{50} (μM)
1	A		−H	−H	2.0
2	A	−H		−H	1.2
3	A	−H	−H		2.0
4	B				9.8
5	A		−H	−H	2.0
6[a]	A	−H		−H	0.9
7	A	−H	−H		2.2
8	B				8.5
9	A		−H	−H	1.8
10	A		−H	−H	2.1
11	A		−H	−H	2.0

Table 1 (*continued*)

No.	Form	R_1	R_2	R_3	GC_{50} (μM)
12	A	−H		−H	0.8
13	A	−H		−H	1.6
14	A	−H		−H	1.0
15[a]	A	−H	−H		2.0
16	A	−H	−H		2.0
17	A	−H	−H		2.0
18	B				9.0

(*continued on next page*)

Table 1 (*continued*)

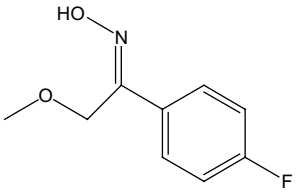
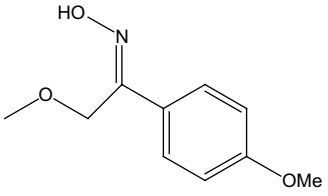
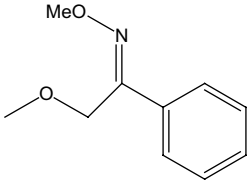
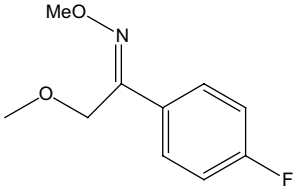
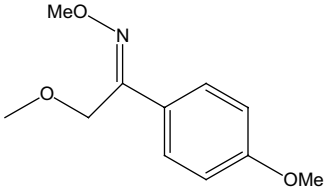
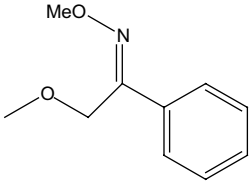
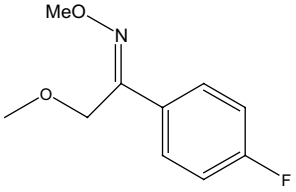
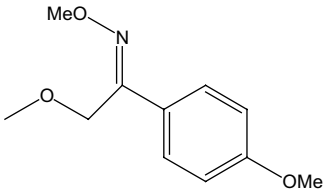
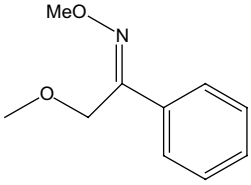
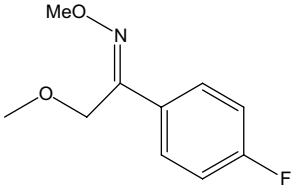
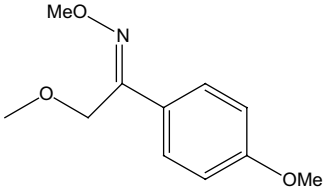
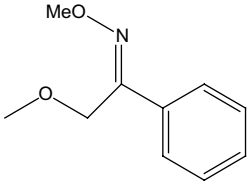
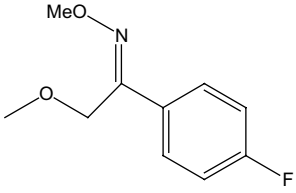
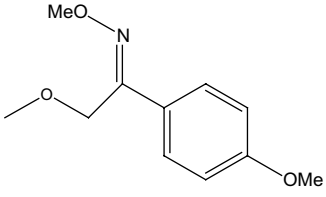
No.	Form	R_1	R_2	R_3	GC_{50} (μM)
19[a]	B				7.8
20	B				7.6
21[a]	A		−H	−H	1.6
22	A		−H	−H	2.0
23	A		−H	−H	2.0
24	A	−H		−H	2.4
25[a]	A	−H		−H	2.3

Table 1 (*continued*)

No.	Form	R_1	R_2	R_3	GC_{50} (μM)
26	A	−H		−H	2.0
27	A	−H	−H		6.6
28	A	−H	−H		2.7
29	A	−H	−H		2.5
30	B				8.2
31	B				6.4
32[a]	B				7.3

[a] Compounds in the test set.

Table 2
Calculational results using the QSAR models

No.	Form	E_{LUMO} (a.u.)	Q_{C6} (a.u.)	pGC_{50} (expt.)	Reg. dev.[a] for Eq. (1)	Pred. dev.[b] for Eq. (1)	q^2_{n-i}[c]	Reg. dev.[a] for Eq. (2)	Pred. dev.[b] for Eq. (2)
1	A	−0.0747	−0.2506	5.699	0.05	0.06	0.778	0.07	0.08
2	A	−0.0728	0.2997	5.921	−0.05	−0.06	0.760	−0.05	−0.06
3	A	−0.0706	−0.2764	5.699	−0.14	−0.14	0.787	−0.12	−0.12
4	B	−0.0607	−0.2758	5.009	0.13	0.15	0.764	0.14	0.17
5	A	−0.0740	−0.2507	5.699	0.02	0.02	0.758	0.04	0.04
7	A	−0.0699	−0.2764	5.658	−0.12	−0.13	0.773	−0.11	−0.11
8	B	−0.0599	−0.2758	5.071	0.03	0.04	0.733	0.05	0.06
9	A	−0.0734	−0.2510	5.745	−0.05	−0.06	0.765	−0.03	−0.04
10	A	−0.0760	−0.2504	5.678	0.13	0.14	0.774	0.15	0.17
11	A	−0.0734	−0.2511	5.699	−0.01	−0.01	0.764	0.01	0.01
12	A	−0.0721	0.2986	6.097	−0.26	−0.31	0.772	−0.26	−0.31
13	A	−0.0742	0.2973	5.796	0.13	0.16	0.774	0.13	0.16
14	A	−0.0719	0.2989	6.000	−0.17	−0.20	0.774	−0.17	−0.20
16	A	−0.0715	−0.2763	5.699	−0.10	−0.10	0.783	−0.08	−0.09
17	A	−0.0694	−0.2764	5.699	−0.19	−0.20	0.795	−0.17	−0.18
18	B	−0.0594	−0.2754	5.046	0.04	0.04	0.751	0.05	0.06
20	B	−0.0592	−0.2758	5.119	−0.04	−0.05	0.760	−0.03	−0.04
22	A	−0.0753	−0.2505	5.699	0.08	0.09	0.780	0.10	0.11
23	A	−0.0730	−0.2513	5.699	−0.02	−0.03	0.778	0.00	−0.01
24	A	−0.0718	0.2991	5.620	0.21	0.25	0.805	0.21	0.25
26	A	−0.0716	0.2992	5.699	0.12	0.15	0.781	0.12	0.14
27	A	−0.0692	−0.2760	5.180	*0.32*	*0.34*	*0.818*	–	–
28	A	−0.0712	−0.2764	5.569	0.02	0.02	0.779	0.04	0.04
29	A	−0.0692	−0.2764	5.602	−0.10	−0.10	0.784	−0.08	−0.09
30	B	−0.0589	−0.2754	5.086	−0.02	−0.03	0.755	−0.01	−0.01
31	B	−0.0617	−0.2758	5.194	−0.01	−0.01	0.765	0.00	0.00

[a] The regression (reg.) deviations (dev.) for all compounds, reg. dev. = pGC_{50} (reg.) − pGC_{50} (expt.).
[b] The prediction (pred.) deviation (dev.) for compound i ($i = 1 \ldots n$), pred. dev. = pGC_{50} (pred.)$_i$ − pGC_{50} (expt.)$_i$, here pGC_{50} (pred.)$_i$ is calculated by the reg. equation obtained after leaving compound i out.
[c] The square of LOO cross-validation coefficient of compound i is computed by the new cross-validation procedure after leaving this datum point (no. i) out from n compounds.

3.2. Discussion on the selected two parameters

It is very interesting that this QSAR model Eq. (2) only has two descriptors, moreover, they belong to quantum-chemical descriptors, that is to say, the energy of the lowest unoccupied molecular orbital (E_{LUMO}) and the net charges of C atom at site 6 on aromatic rings (Q_{C6}). These two descriptors are most responsible for the cytotoxicity and this model shows that decreasing E_{LUMO} value and increasing Q_{C6} value can lead to an increase in cytotoxicity against HeLa.

The evaluation of the correlation coefficient of E_{LUMO} versus pGC_{50} ($R = -0.894$) shows that E_{LUMO} has the higher contribution to the pGC_{50}; it implies that the frontier MO-controlled process plays a crucial role in the interaction of the compound with biomacromolecule. Thus, decreasing the energy of the LUMO orbital can greatly increase the activity of the compound. From Table 1, we can easily find that the flavone derivatives have considerably higher activities than the isoflavone derivatives because they have much lower E_{LUMO}. Such a trend can be theoretically explained by the law of polarity alternation [28,29] and the idea of polarity interference [27,30,31].

For the above purpose, we first should draw sketches of polarity interference as Fig. 3, where a solid arrowhead toward a more electronegative atom represents the direction of primary bond polarity; a dotted arrowhead represents the direction of induced polarity. To make a schematic map simple, we just draw the dotted arrowheads on the bonds formed between α-site and β-site atoms. Based on the idea of polarity interference, for different isomers, the more the arrowheads with the same direction, the lower the energy of compound is. From Fig. 3, we can

Table 3
Correlation matrix for pGC_{50} and selected parameters in the Eq. (2)

	pGC_{50}	E_{LUMO}	Q_{C6}
pGC_{50}	1.000	–	–
E_{LUMO}	−0.894	1.000	–
Q_{C6}	0.535	−0.328	1.000

Table 4
Structures and computational results for six compounds in the test set

No.	Form	E_{LUMO} (a.u.)	Q_{C6} (a.u.)	pGC_{50} (expt.)	pGC_{50} (calc.)[a]	Dev.[b]
6	A	−0.0724	0.3000	6.046	5.854	−0.19
15	A	−0.0695	−0.2760	5.699	5.535	−0.16
19	B	−0.0634	−0.2757	5.108	5.271	0.16
21	A	−0.0728	−0.2511	5.796	5.686	−0.11
25	A	−0.0739	0.2977	5.638	5.918	0.28
32	B	−0.0587	−0.2758	5.137	5.067	−0.07

[a] Predictive activities were calculated using the QSAR model.
[b] Dev. = pGC_{50} (calc.) − pGC_{50} (expt.).

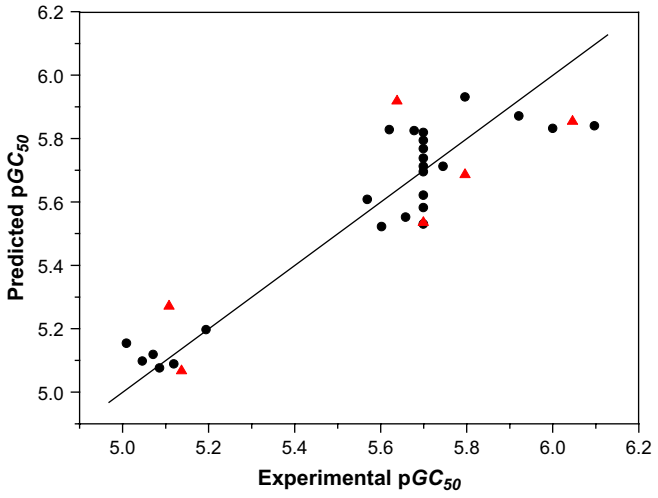


Fig. 2. Plot of predicted activities versus experimental ones for the QSAR model Eq. (2), in which 25 compounds are the training set (dots) and correspondingly 6 compounds are the test set (triangles).

observe that flavone only has same-directional arrowheads, while isoflavone not only has same-directional arrowheads, but also has reverse-directional arrowheads. So we can predict that the total energy of flavone is lower than that of isoflavone, and correspondingly the E_{LUMO} of flavone is generally lower than that of isoflavone (see Table 2).

According to the molecular orbital theory of chemical reactivity, the lower energy of the LUMO means the compound more easily accepts electron in chemical reaction and thus plays a role as an electrophilic reagent. Therefore, we can reasonably speculate that these compounds act as election acceptors in the interaction with drug-receptor, and electron transfer happens from the receptor to the compound. It is well-known that many antitumor drugs exert their action by binding to DNA [32–34], and that the interaction between them is essentially via orbital interaction or charge transfer process in binding of the compound to DNA in intercalative model. Moreover, many theoretical studies have shown that the bases and base-pairs of DNA are electron donors, because the HOMO energy of DNA-base-pairs is rather high (−1.27 eV) [33], whereas the intercalative compound are electron-acceptors because their LUMO energies are negative and quite low.

The calculations show that the LUMO energies of these compounds are rather low (−2.07 to −0.160 eV) and all of them are lower than the HOMO energy of DNA-base-pairs (−1.27 eV). In addition, the main part (three aromatic rings)

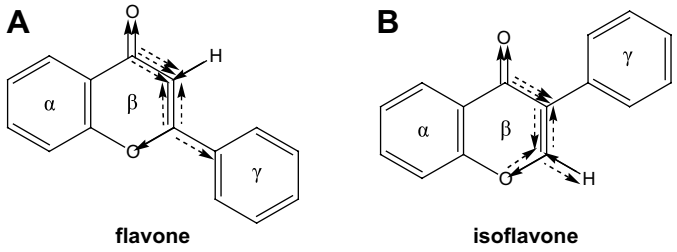


Fig. 3. Schematic maps of polarity interference for (A) flavone and (B) isoflavone.

of these compounds are conjugative, (see Fig. 1) and thus they have an excellent plane that can intercalate between DNA-base-pairs in an intercalative mode. Hereby, the lower the LUMO energy, the stronger the DNA-binding affinity, and thus the higher the activity of the compound is. Therefore, based on this action mechanism, the result that E_{LUMO} is selected as a very important descriptor in the QSAR model should be reasonable.

In order to explain the action mechanism, the spatial LUMO distributions of some compounds as examples were drawn and shown in Fig. 4. From Fig. 4, we can clearly see that the LUMOs of these compounds are all distributed on the main conjugative part (composed by three aromatic rings α, β and γ). We can further find that, the LUMOs of compound **12**, **22** and **29**, which are flavone derivatives, are uniformly distributed on the three aromatic rings, in particular, on ring γ. However, the LUMO of compound **4**, which is one of isoflavone derivatives, is only distributed on the two aromatic rings (α and β). The similar LUMO distribution characteristics are also found for other flavone and isoflavone derivatives. That is to say, the components of LUMO of the compound with a stronger cytotoxicity against HeLa must be distributed on the three aromatic rings, especially on phenyl ring γ. Consequently, we can suggest that the three aromatic rings are the active parts in the interaction between the compound and DNA-base-pairs, and the LUMO distribution on ring γ seems to be most important because these compounds can be assumed to intercalate between DNA-base-pairs along the direct of ring γ–β–α due to a big non-conjugative substituent on ring α or β.

On the other hand, the electronic parameter Q_{C6} exhibits positive correlation coefficient with pGC_{50} ($R = 0.535$), so the higher the Q_{C6} value, the higher the activity of the compound is. Such regularity can be easily found through comparisons between compounds **4** and **12**, **14** and **18**, and other ones. The charges on the atom C6 may be understood as a representation of the total charges of the conjugative part of the compound. Since the more the total positive charge of the conjugative part, the stronger the interaction between the conjugative part and DNA-base-pairs, because the former is electron-acceptor and the latter is electron-donator. We also find that the Q_{C6} values of the isoflavone derivatives are all negative and their negative values are rather great (ca. −0.2755 a.u.), so their cytotoxicities are correspondingly smaller; on the contrary, the Q_{C6} values of flavone derivatives are most positive and rather great, thereby their cytotoxicities are correspondingly higher.

On the basis of above analysis, we can clearly observe that the two selected parameters (E_{LUMO} and Q_{C6}) in the QSAR model are all disadvantageous to the cytotocicity of isoflavone derivatives against HeLa. Therefore, the established QSAR successfully explains a very important experimental result or regularity, i.e., flavone derivatives have considerably higher activities than isoflavone derivatives.

It is notable that the parameters about oxime- or methyloxime-containing substituents do not give any significant correlations with cytotoxicity of these compounds. This means

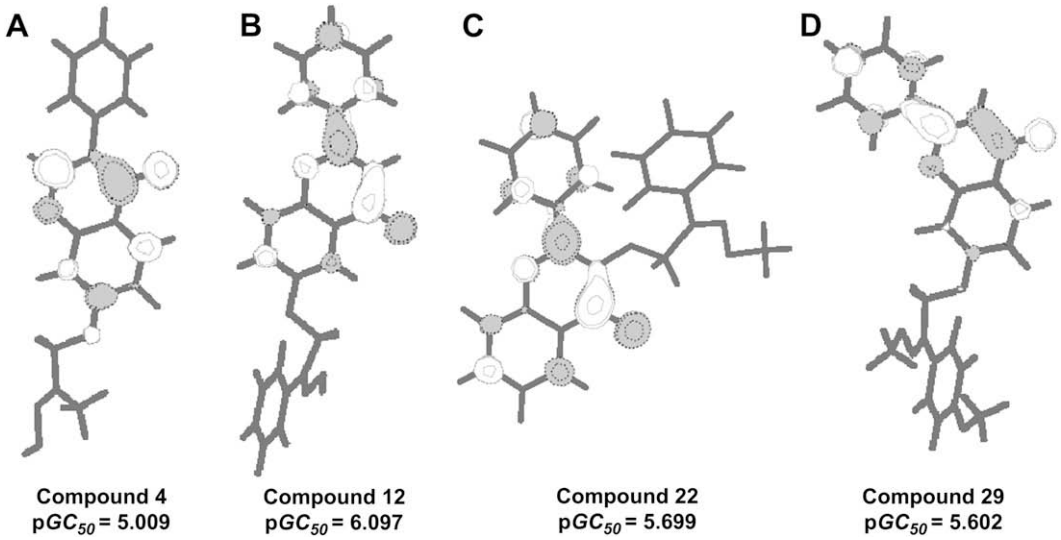


Fig. 4. Spatial LUMO distributions of some compounds.

that they do not carry critical information on cytotoxicity. It may relate to the fact that oxime- and methyloxime-containing substituents play a small role in an intercalative mode because they do not intercalate between DNA-base-pairs and thus result in little influence for cytotoxicity of studied compounds.

In fact, here the two selected descriptors (E_{LUMO} and Q_{C6}) describe the molecular inherent electronic properties which highly influence the biological activities of molecules. Such descriptors can be easily obtained from the DFT computations, their physical meanings are very clear, and they are easily controlled, thus they can be effectively used in the action mechanism analysis and molecular design. Based on the above established model, the molecular design of the compounds with higher anticancer activity is presented.

3.3. Design of new compounds with higher anticancer activity

According to the above discussion, we can clearly see that E_{LUMO} plays a critical role affecting the cytotoxicity of the studied compounds against HeLa and these compounds possibly act as intercalative agents in interaction with DNA. Thus in order to design new compounds with higher cytotoxicity against HeLa, we should emphatically consider decreasing the E_{LUMO} value as well as increasing the planarity and plane area of compound, because they greatly affect the DNA-binding affinity of the compound. Moreover, we should only consider the design of flavone derivatives because they have considerably higher activities than isoflavone derivatives.

We carried out structural modification starting from compound **12** as template (see Fig. 7), because it has the highest cytotocicity. Since the substituents on compound **12** are appropriate, the emphasis of the molecular design will be focused on the active three aromatic rings, especially on phenyl ring γ. Moreover, since the substitution on the 2′-position or 6′-position of the phenyl ring γ may increase hindrance for the DNA-binding in intercalative mode due to the chromophore incoplanarity, the other sites on the phenyl ring γ can be considered [35]. Therefore, we only need to emphatically consider the substituents on 3′,4′-positions or 5′-position of phenyl ring γ in order to improve the cytotoxicity of compound.

In addition, we consider that the substitution of an electron-withdrawing group can make the LUMO energy reduce, whereas the substitution of an electron-donating group can make the LUMO energy increase, and that electron-withdrawing groups located on the position of same-directional interference (alternative positions) can considerably reduce the LUMO energy based on the law of polarity alternation and the idea of polarity interference (see Fig. 5). In particular, we also consider that nitro group has some excellent performance as follows: (1) It not only has high electron-withdrawing ability but also can conjugate with phenyl ring γ, resulting in enlargement of plane area as well as supplying more active sides. (2) It can result in intramolecular interaction with *ortho* hydroxy or amido group via intramolecular hydrogen bond to form planar six cycle (Fig. 6).

According to the above established QSAR model and considerations, 5 new compounds with higher cytotoxic activity against HeLa have been theoretically designed. The structures of the 5 designed compounds and their two parameter values calculated by the same methods, as well as the pGC_{50} values predicted by the QSAR model are listed in Table 5.

From Table 5, we can find that the cytotoxicities of the designed 5 new compounds (**D1**–**D5**) are all much higher

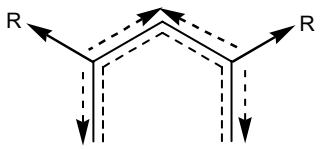


Fig. 5. Schematic map of polarity interference (R = electron-withdrawing group).

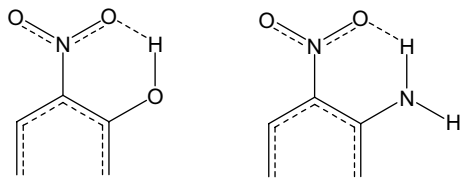


Fig. 6. Representation of intramolecular hydrogen bonding.

($pGC_{50} = 7.7-8.7$), compared with anyone of the 32 compounds (**1**–**32**) in the training set and test set, because their E_{LUMO} values are rather lower (from −0.14 to −0.12 a.u.) and the Q_{C6} values keep an rather large constant (~0.30 a.u.). Such results further suggest that this model has a strong predictive ability and can be prospectively used in molecular design or structural modification.

The action mechanism and design of compounds with cytotoxicity remain complicated, and many factors might affect their biological activities, e.g., differences of uptake and transport into cells and other physicochemical properties [36,37], etc. Here, we just offer some theoretical considerations based on the QSAR research and expect these results to be useful references for the experimental works.

4. Conclusions

A QSAR of a series of lavone and isoflavone derivatives with cytotoxicity against HeLa has been studied by using DFT and MM2 calculations as well as statistical methods. It is very interesting to find that the cytotoxicity of this kind of compound only depend on two quantum-chemical descriptors: the LUMO energy (E_{LUMO}) and the atomic net charges on the C6 (Q_{C6}). The established optimal QSAR shows not only acceptable regressive performance ($R^2_A = 0.852$) but also considerable predictive ability ($q^2 = 0.818$). Moreover, the predictive correlation coefficient R^2_{pred} for the test set reaches 0.738, further showing that the established QSAR has excellent predictive ability. Based on this QSAR model, some useful insight into understanding the action mechanism at molecular level are obtained as follows: (1) These compounds can act as election acceptors in the interaction with DNA-base-pairs. (2) A very interesting experimental result or regularity, i.e., flavone derivatives have considerably higher activities than isoflavone derivatives, has been reasonably explained. In addition, 5 new compounds with higher cytotoxicity have been theoretically designed. Such theoretical results can be used as references for the further experimental works.

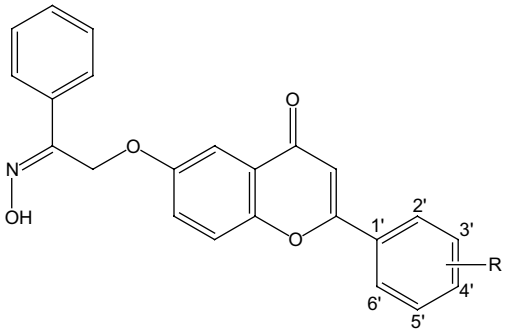


Fig. 7. Structural schematic diagram of the designed compounds.

Table 5
Structures and computational results for the 5 designed compounds

Comp.	R	E_{LUMO} (a.u.)	Q_{C6} (a.u.)	pGC_{50} (calc.)
D1	3′-NO_2	−0.1154	0.3021	7.700
D2	4′-NO_2	−0.1217	0.3030	7.970
D3	3′-NO_2, 5′-NO_2	−0.1378	0.3052	8.662
D4	4′-NO_2, 3′-OH	−0.1309	0.3033	8.365
D5	3′-NO_2, 4′-NH_2, 5′-NO_2	−0.1399	0.3030	8.751

Acknowledgements

The financial supports of the National Natural Science Foundation of China and the Natural Science Foundation of Guangdong Province of China are gratefully acknowledged.

References

[1] D. Arthan, J. Svasti, P. Kittakoop, D. Pittayakhachonwut, M. Tanticharoen, Y. Thebtaranonth, Phytochemistry 59 (2002) 459–463.
[2] B.F. Rasulev, N.D. Abdullaev, V.N. Syrov, J. Leszczynski, QSAR Comb. Sci. 24 (2005) 1056–1065.
[3] V.C. Blank, C. Poli, M. Marder, L.P. Roguin, Bioorg. Med. Chem. Lett. 14 (2004) 133–136.
[4] J. Lameira, C.N. Alves, V. Moliner, E. Silla, Eur. J. Med. Chem. 41 (2006) 616–623.
[5] E. Middleton Jr., C. Kandaswami, T.C. Theoharides, Pharmacol. Rev. 52 (2000) 673–751.
[6] J.A. Beutler, J.H. Cardellina, C.M. Lin, E. Hamel, G.M. Cragg, M.R. Boyd, Bioorg. Med. Chem. Lett. 3 (1993) 581–584.
[7] H.J. Woerdenbag, I. Merfort, C.M. Passreiter, T.J. Schmidt, G. Willuhn, W. van Uden, N. Pras, H.H. Kampinga, A.W.T. Konings, Planta Med. 60 (1994) 434–437.
[8] C. Pouget, F. Lauthier, A. Simon, C. Fagnere, J.P. Basly, C. Delage, A.J. Chulia, Bioorg. Med. Chem. Lett. 11 (2001) 3095–3097.
[9] Y.C. Chen, S.C. Shen, J.M. Chow, C.H. Ko, S.W. Tseng, Int. J. Oncol. 25 (2004) 661–670.
[10] T.C. Wang, I.L. Chen, P.J. Lu, C.H. Wong, C.H. Liao, K.C. Tsiao, K.M. Chang, Y.L. Chen, C.C. Tzeng, Bioorg. Med. Chem. 13 (2005) 6045–6053.
[11] T.C. Wang, I.L. Chen, C.M. Liu, D.H. Kuo, C.H. Liao, Chem. Biodivers. 2 (2005) 253–263.
[12] O. Isayev, B. Rasulev, L. Gorb, J. Leszczynski, Mol. Divers. 10 (2006) 233–245.
[13] R. dos Santos, C.A. Kuhnen, R.A. Yunes, Arch. Pharm. Chem. Life Sci. 339 (2006) 227–237.
[14] F.A. Pasha, H.K. Srivastava, P.P. Singh, Bioorg. Med. Chem. 13 (2005) 6823–6829.
[15] J.C. Chen, L. Qian, W.J. Wu, L.M. Chen, K.C. Zheng, J. Mol. Struct. THEOCHEM 756 (2005) 167–172.
[16] CambridgeSoft corp., 100 Cambrige Park, MA 02140–2317, USA.
[17] A.D. Becke, J. Chem. Phys. 98 (1993) 1372–1377.
[18] J.E.Carpenter,F.Weinhold,J.Mol.Struct.THEOCHEM169(1988)41–46.
[19] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo,

J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2005.

[20] HyperChem, Release 6.03, Hypercube Inc., 2001.http://www.hyper.com/.

[21] SPSS vs 9.0; SPSS Inc.: Chicago, IL, 1999.

[22] K. De, C. Sengupta, K. Roy, Bioorg. Med. Chem. 12 (2004) 3323–3332.

[23] I.V. Tetko, V.Y. Tanchuk, A.E. Villa, J. Chem. Inf. Comput. Sci. 41 (2001) 1407–1421.

[24] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.

[25] D.R. Roy, U. Sarkar, P.K. Chattaraj, A. Mitra, J. Padmanabhan, R. Parthasarathi, V. Subramanian, S. Van Damme, P. Bultinck, Mol. Divers. 10 (2006) 119–131.

[26] S. Zhang, L. Wei, K. Bastow, W. Zheng, A. Brossi, K.H. Lee, A. Tropsha, J. Comput. Aided Mol. Des. 21 (2007) 97–112.

[27] J.C. Chen, Y. Shen, S.Y. Liao, L.M. Chen, K.C. Zheng, Int. J. Quantum Chem. 107 (2006) 1468–1478.

[28] J.A. Pople, M.S. Gordon, J. Am. Chem. Soc. 89 (1967) 4253–4261.

[29] S. Fliszar, J. Am. Chem. Soc. 94 (1972) 1068–1073.

[30] K.C. Zheng, F. He, Z.T. Xu, F.C. Yun, Acta Phys.-Chim. Sin. 15 (1999) 698–702 [in Chinese].

[31] K.C. Zheng, J.P. Wang, W.L. Peng, X.W. Liu, F.C. Yun, J. Phys. Chem. A 105 (2001) 10899–10905.

[32] D. Řeha, M. Kabeláč, F. Ryjáček, J. Šponer, J.E. Šponer, M. Elstner, S. Suhai, P. Hobza, J. Am. Chem. Soc. 124 (2002) 3366–3376.

[33] N. Kurita, K. Kobayashi, Comput. Chem. 24 (2000) 351–357.

[34] J. Li, L.C. Xu, J.C. Chen, K.C. Zheng, L.N. Ji, J. Phys. Chem. A 110 (2006) 8174–8180.

[35] G.J.Atwell,B.C.Baguley,W.A.Denny,J.Med.Chem.32(1989)396–401.

[36] A.C.G. Hotze, S.E. Caspers, D. de Vos, H. Kooijman, A.L. Spek, A. Flamigni, M. Bacac, G. Sava, J.G. Haasnoot, J. Reedijk, J. Biol. Inorg. Chem. 9 (2004) 354–364.

[37] A.H. Velders, K. Van der Schilden, A.C.G. Hotze, J. Reedijk, H. Kooijman, A.L. Spek, J. Chem. Soc. Dalton Trans. (2004) 448–455.