Syntheses and QSAR Studies of Benzylimidazole Derivatives and Benzylcarbazole as Potential Aromatase Inhibitors

Y. Dai[1,*], Y. Xiao[1], Q. Wang[1], S. Wei[1], X. Zhang[1], Z. Ma[2], H. Zheng[3], M. Hou[1] and T. Zhang[1]

[1]Key Laboratory of Industrial Fermentation Microbiology (Tianjin University of Science & Technology), Ministry of Education, College of Bioengineering, Tianjin University of Science and Technology, Tianjin 300457, P.R. China
[2]Department of Ocean Science and Engineering, Zhejiang University, Hangzhou 310058, P.R. China
[3]School of Life Science and Technology, China Pharmaceutical University, Nanjing 210009, P.R. China

*Corresponding author: Fax: +86 22 60602298; Tel: +86 22 60601265; E-mail: yjdai@126.com

In this study, in order to explore new structures and chemical entities as aromatase inhibitors, benzylcarbazole, 12 benzylimidazole derivatives with different substituents on both phenyl and imidazole rings were synthesized and their aromatase inhibitory were evaluated with fluorescent substrate detection method. The results showed that the compounds with carboxyl and ester groups in phenyl ring show better inhibitory activity. The introduction of alkyl groups in imidazole may improve the aromatase inhibitory activity. Most of the compounds were more potent than aminoglutethimide and tamoxifen. 2-[2-{(2-ethyl-4-methyl-1*H*-imidazol-1-yl)methyl}phenyl]acetic acid and benzylcarbazole have the highest bioactivities with IC_{50} values of 6.19 μM and 2.72 μM respectively. A meaningful QSAR model with LOF = 0.00359, R^2 = 0.9914, Adj-R^2 = 0.9853, R^2cv = 0.9639, F = 161.8, was constructed with genetic functional algorithm using discovery studio 2.1 package.

Keywords: Benzylimidazole, Benzylcarbazole, Aromatase inhibitors, QSAR.

INTRODUCTION

Cancer is the leading cause of death among women between the ages of 30 and 54, with breast and uterine cancers comprising 10 % and 28 %, respectively, of all cancers in females per year[1]. Excluding skin cancers, breast cancer is the most common malignancy among women, accounting for nearly 1 in 3 cancers diagnosed among women in the United States and it is the second leading cause of cancer death among women[2]. Although the rate of postmenopausal breast cancer were ten fold higher in Americans than in Chinese, whereas the rates for Chinese-Americans were intermediate[3], compared with 2000 years. There are 470 thousands more new breast cancer cases and 130 thousands more deaths from breast cancer in 2005 in China[4]. Approximately 2 in 3 of postmenopausal breast cancer patients have estrogen-dependent breast cancer, which includes estrogen receptors and requires estrogen for tumor growth. Estradiol is the most potent endogenous estrogen. Consequently, inhibition of estrogen biosynthesis by means of selective aromatase inhibitors is a potentially useful thera-peutic option in hormone-sensitive breast cancer[5]. Estradiol is biosynthesized from androgens by the cytochrome P450 enzyme called "aromatase"(CYP19)[6], with the highest levels

of enzyme present in the ovaries of premenopausal women, in the placenta of pregnant women and in the peripheral adipose tissues of postmenopausal women and of men[7].

Furthermore, expression of aromatase is the highest in or near breast tumor sites[8,9]. Thus, aromatase is a key drug target for the treatment of breast cancer. Aromatase inhibitors have emerged over the last many years as modulators of the growth-stimulatory effects of estrogens in estrogen-dependent breast cancer[10]. Aromatase inhibitors are divided into two categories: steroidal aromatase inhibitors and nonsteroidal aromatase inhibitors. Steroidal inhibitors have been developed to now, such as exemestane, build upon the basic androstenedione nucleus and incorporate substituents at varying positions on the steroid[11]. Most of aromatase inhibitors belong to nonsteroidal aromatase inhibitors, such as tamoxifen (which has been used as the standard anti-estrogen in treating such tumors[12]), aminoglutethimide, anastrozole, letrozole, *etc.* **(Scheme-I)**. Now anastrozole and letrozole are approved aromatase inhibitors for the treatment of metastatic estrogen-dependent breast cancer[13-17]. In spite of their clinical success, negative side effects and partial selectivity of existing aromatase inhibi-tors are significant issues which must be addressed. Aromatase inhibitors are associated with osteoporosis, reproductive
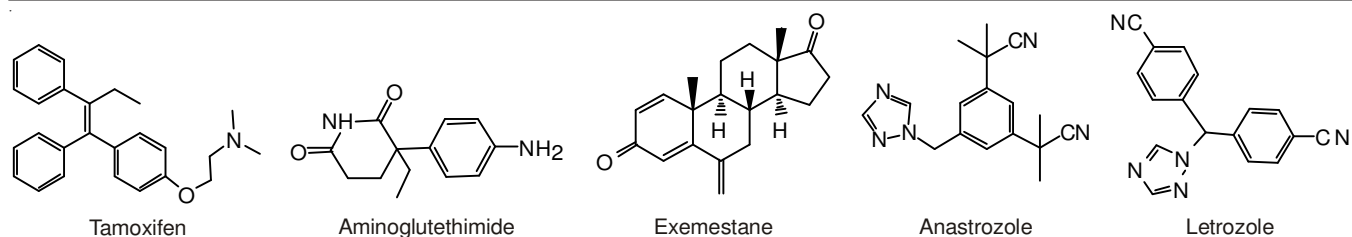
Scheme-I: Commonly used drugs for the treatment of breast cancer

problems and androgenic side effects. These factors in combination necessitate the development of more selective new aromatase inhibitors for the treatment of breast cancer and up to now the search for potent and selective aromatase inhibitors still remains an attractive subject[18], Moreover, alternative strategies such as development of single agents against multiple drug targets[19,20] are being evaluated from different research groups to deal with the complexity and multiplicity of factors involved in the development of hormone-dependent breast cancer. Various techniques were applied to search for the new aromatase inhibitors, including using the existing drugs as templates[21], structural modification of natural products[22], high-throughput docking[23] or structure-guided drug design[24], *etc*.

It was found that most of the non-steroidal aromatase inhibitors are composed of 3 or 4 aromatic rings[25]. However, some compounds with simple structure and comparatively fewer aromatic rings also behave good aromatase inhibitory activities. It has been reported that inhibitory activities (IC_{50}) of CHEMBL 349822 (with 2 aromatic rings) toward human aromatase enzyme was 0.368 µM[26]. These aromatase inhibitors have advantage of higher activity, relatively simple structure and easier to synthesize and then we were interested in one group of these aromatase inhibitors **(Scheme-II)**. In particular, the stereochemically defined functionalized phenyl ring is essential for potent activity, so researchers added some substituents to the phenyl ring to improve the activity of these aromatase inhibitors[26-28]. In this research, we intend to change the structure of imidazole ring (such as adding some substituents to the imidazole ring, or using carbazole ring instead) in order to obtain the new aromatase inhibitors with high activity and 13 compounds were obtained **(Scheme-III)**.

Molecular modeling has become a useful tool for the characterization of structure-function relationships for a receptor and for the identification of structural motives of both ligands and receptors, which may play important roles in the binding processes[30]. Quantitative structure-activity relationship (QSAR) is an important tool to keep the number of synthesized and tested compounds at a minimum in the process of development of new drugs[31]. Nowadays, three-dimensional (3D) quantitative structure activity relationship (3D-QSAR) techniques, such as comparative molecular field analysis (CoMFA) are

routinely used in modern drug design to help in understanding drug-receptor interaction. Relatively large number of quantitative structure-activity relationship (QSAR) studies involving inhibitors of dipeptidyl peptidase IV, glycogen phosphorylase, acetylcholinesterase, HIV protease, estrogens and so on have been reported[32-37]. Also there are some 2D-3D hybrid QSAR studies on the steroid and non-steroid aromatase inhibitors have been conducted with molecular shape analysis descriptors along with thermodynamic and structural descriptors and also with selected topological parameters on structurally diverse datasets of aromatase inhibitors[38]. Since Ghosh *et al.*[39] published the crystal structure of CYP19 (Fig. 1), various ligand-receptor interaction investigations between CYP19 and the small molecule inhibitors using docking process were reported. These computational techniques have been proved particularly helpful in the design of novel, more potent inhibitors by revealing the mechanism of drug-receptor interaction[40].
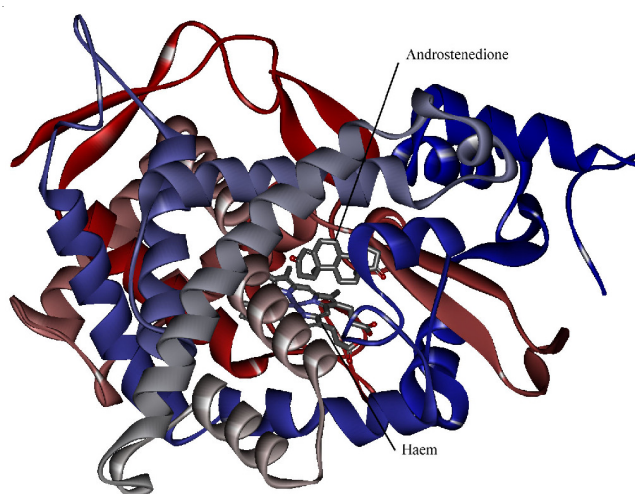

Fig. 1

In this study, 13 compounds which contain 12 benzyl imidazoles derivatives and benzyl carbazole were synthesized and their aromatase inhibitory activities were determined with molecular fluorescence analysis method[41]. A QSAR model was built on the basis of the synthesized compounds and some
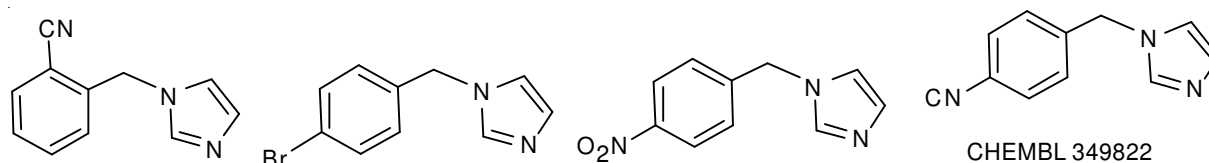
Scheme-II: Some aromatase inhibitors with the structure of benzyl-imidazoles 26, 29
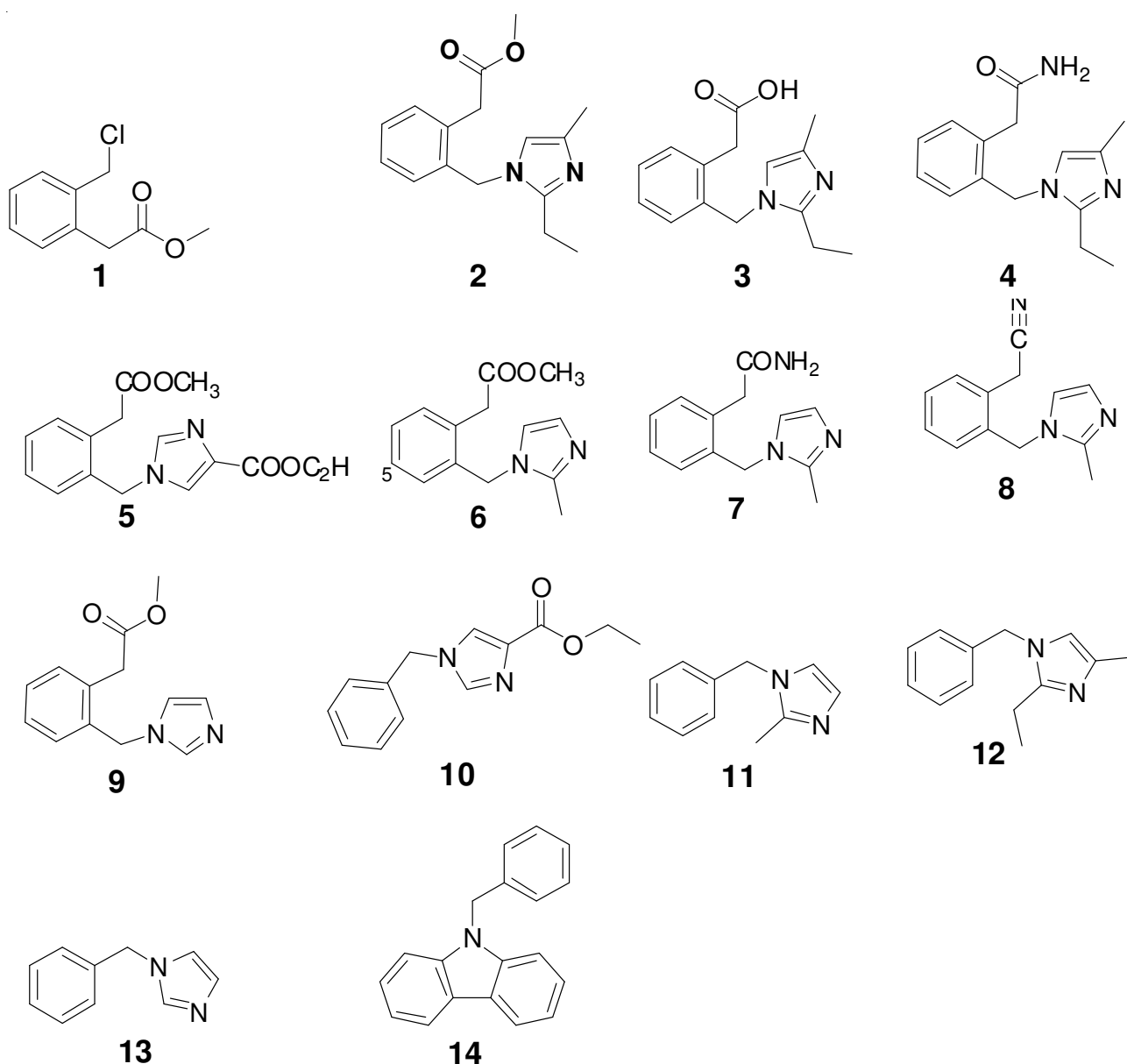
Scheme-III: Aromatase inhibitors synthesized in this paper

reported benzyl imidazole derivatives, also molecular docking for some ligands with higher bioactivities to the CYP19 binding site was conducted.

EXPERIMENTAL

It is well known that benzyl imidazole skeleton was easily obtained by the reaction of benzyl halide and imidazole at room temperature and N-substituted imidazole was easily obtained by the reaction of halogenated hydrocarbons and imidazole or carbazole. In order to avoid the substitution reaction taken place at the *para*-position of the benzene ring, 3-isochromanone was adopted as the source regent to prepare the *ortho*-substituted benzyl halide (compound **1**). All reaction for the object compounds are not only easy to operate, but also have less byproducts, which makes it easy for the purification of the products.

Synthesis of compounds 2-9: The lactonic ring of 3-isochromanone was opened with SOCl$_2$ under basic conditions

to give the phenyl acetate derivative (compound **1**) in 85.67 % yield[42], which was reacted further with imidazole, carbazole in CHCl$_3$ under K$_2$CO$_3$ basic condition with 10 h reflux to form compound **9** in 73.22 % yield. In a similar fashion, compound **2** was obtained from compound **1** *via* the K$_2$CO$_3$ in CHCl$_3$ by reaction with 2-ethyl-4-methyl-1*H*-imidazole. The primary ester group in compound **2** was converted to amide **4** *via* nucleophilic substitution with NH$_3$ using ammonia water at room temperature to give a yield of 53.67 %. Compound **3** was obtained by hydrolysis of ester bond of compound **2**. Similarly, compound **1** was also converted to compound **5** in 67.55 % yield by reaction with 4-imidazole carboxylate. The substitution of the chlorine group of compound **1** with 2-methyl imidazole led to the formation of compound **6** and compound **7** was obtained by ammonia substitution of the ester group in compound **6**. The nitrile group of compound **8** was formed by the dehydration reaction of compound **7** with POCl$_3$ in ethyl acetate in 52.82 % yield.

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 2: δ1.105 (t, 3H, *J* = 7.6 Hz, CH₂C**H₃**), δ2.063 (s, 3H, imidazole-CH₃), δ2.490 (m, 2H, **CH₂**CH₃), δ3.623 (s, 3H, OCH₃), δ3.802 (s, 2H, CH₂CO), δ5.086 (s, 2H, imidazole-CH₂), δ6.41 (s, 1H, imidazole**H**), δ6.632 (m, 2H, ArH), δ7.237 (m, 2H, ArH).

¹H NMR (D₂O, ppm, 400 MHz) of compound 3: δ1.100 (t, 3H, *J* = 7.6 Hz, CH₂C**H₃**), δ2.083 (s, 3H, imidazole-CH₃), δ2.536 (m, 2H, **CH₂**CH₃), δ3.503 (s, 2H, CH₂CO), δ4.703 (s, 2H, imidazole-CH₂), δ6.556(s, 1H, imidazole**H**), d7.216 (m, 4H, ArH).

¹H NMR (DMSO, ppm, relative to TMS, 400 MHz) of compound 4: δ1.105 (t, 3H, *J* = 7.6 Hz, CH₂C**H₃**), δ2.058 (s, 3H, imidazole-CH₃), δ2.504 (m, 2H, **CH₂**CH₃), δ3.497 (s, 2H, CH₂CO), δ5.164 (s, 2H, imidazole-CH₂), δ6.514 (d, 1H, *J* = 7.2 Hz, imidazole**H**), δ6.643 (d, 1H, *J* = 0.8 Hz ArH), δ6.985 (s, 1H, NH), δ7.234 (m, 3H, ArH + NH), δ7.538(s, 1H, ArH).

¹H NMR (DMSO, ppm, 400 MHz) of compound 5: δ1.169 (t, 3H, OCH₂C**H₃**), δ3.628 (s, 3H, COOCH₃), δ3.869 (s, 2H,COO**CH₂**Ar), δ4.171 (m, 2H, imidazole-COO**CH₂**), δ5.555 (s, 2H, imidazole-**CH₂**Ar), δ6.559 (m, 1H, imidazole**H**), δ7.247 (m, 3H, ArH), δ7.740 (δ, 2H, *J* = 0.8 Hz, ArH), δ7.990 (s, 1H, imidazole**H**).

¹H NMR (DMSO, ppm, 400 MHz) of compound 6: δ2.169 (s, 3H, imidazole-**CH₃**), δ3.627 (s, 3H, COOCH₃), δ3.824 (s, 2H, Ar**CH₂**CO), δ5.176 (s, 2H, Ar**CH₂**-imidazole), δ6.554 (d, 1H, *J* = 6.8 Hz, imidazole**H**), δ6.813 (d, 1H, *J* = 1.2 Hz, imidazole**H**), δ6.982 (d, 1H, ArH), δ7.247 (m, 3H, ArH).

¹H NMR (DMSO, ppm, 400 MHz) of compound 7: δ2.177 (s, 3H, imidazole-**CH₃**), δ3.516 (s, 2H, Ar**CH₂**CO), δ5.241 (s, 2H, Ar**CH₂**-imidazole), δ6.483 (d, 1H, *J* = 7.2 Hz, imidazole**H**), δ6.804 (d, 1H, 1.2 Hz, imidazole**H**), δ6.993 (d, 2H, *J* = 1.2Hz, ArH), δ7.231(m, 3H, ArH + NH₂), δ7.543 (s, 1H, ArH).

¹H NMR (DMSO, ppm, 400 MHz) of compound 8: δ2.210 (s, 3H, imidazole-**CH₃**), δ4.170 (s, 2H, CNCH₂), δ5.271 (s, 2H, Ar**CH₂**-imidazole), δ6.548 (d, 1H, *J* = 7.6 Hz, imidazole**H**), δ6.887 (s, 1H, imidazole**H**), δ7.067 (s, 1H, ArH), δ7.321 (m, 2H, ArH), δ7.475 (d, 1H, *J* = 7.2 Hz, ArH).

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 9: δ3.621 (s, 2H, Ar**CH₂**CO), δ3.675 (s, 3H, COCH₃), δ5.206 (s, 2H, Ar**CH₂**-imidazole), δ6.864 (s, 1H, imidazole **H**), δ7.013 (d, 1H, *J* = 5.6 Hz, imidazole **H**), δ7.083 (s, 1H, ArH), δ7.292 (m, 3H, ArH), δ7.491 (s, 1H, imidazole **H**).

Synthesis of compounds 10-14: Benzyl chloride was easily reacted with imidazoles at ambient temperature in CHCl₃ under stirring for 24 h. Compounds **10-13** was obtained in the yields of 79.16, 75.56, 75.35 and 73.74 %, respectively. Carbazole derivatives are important pharmaceutical intermediates and recently are concerned as selective androgen receptor modulators[43], however, their aromatase inhibitory activities have not attracted much attention, thus in this article, compound **14** was synthesized by the reaction of carbazole with benzyl chloride in 56.61 % yield.

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 10: δ1.332 (t, 3H, *J* = 7.6 Hz, OCH₂C**H₃**), δ4.292 (m, 2H, **OCH₂**CH₃), δ5.539 (s, 2H, Ar**CH₂**-imidazole), δ7.195 (d, 2H, ArH + imidazole**H**), δ7.338 (m, 3H, ArH), δ7.691 (s, 1H, ArH), δ7.812 (s, 1H, imidazole**H**).

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 11: δ2.349 (s, 3H, imidazole-**CH₃**), δ5.053 (s, 2H, Ar**CH₂**-imidazole), δ6.847 (d, 1H, *J* = 1.2 Hz, imidazole**H**), δ6.966 (d, 1H, *J* = 1.2 Hz, imidazole**H**), δ7.065 (d, 2H, *J* = 6.4 Hz, ArH), δ7.333 (m, 3H, ArH).

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 12: δ1.238 (t, 3H, *J* = 7.6 Hz, imidazole-CH₂C**H₃**), δ2.181 (s, 3H, imidazole-**CH₃**), δ2.595 (m, 2H, imidazole-**CH₂**CH₃), δ4.965 (s, 2H, Ar**CH₂**-imidazole), δ6.508 (s, 1H, imidazole**H**), δ7.041 (d, 2H, *J* = 6.8 Hz, ArH), δ7.304 (m, 3H, ArH).

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 13: δ5.129 (s, 2H, Ar**CH₂**-imidazole), δ6.9261 (s, 1H, imidazole**H**), δ7.173 (m, 3H, Ar + imidazole**H**), δ7.370 (m, 3H, ArH), δ7.560 (s, 1H, imidazole**H**).

¹H NMR (CDCl₃, ppm, relative to TMS, 400 MHz) of compound 14: δ5.557 (s, 2H, Ar**CH₂**-carbazole), δ7.188 (d, 2H, *J* = 6 Hz, ArH), δ7.291 (m, 5H, ArH + carbazole**H**), δ7.404 (d, 2H, *J* = 8 Hz, carbazole**H**), δ7.469 (m, 2H, carbazole H), δ8.175 (d, 2H, *J* = 8 Hz, carbazole**H**).

Inhibitory activities: The inhibitory activities of the target compounds were evaluated by a modification of the method of Luo *et al.*[41].

Molecular docking: The pdb data about the crystal structure of CYP19 complexing with androstenedione (EC: 1.14.14.1, 3EQM.pdb)[39] was obtained from the RCSB protein data bank (http://www.pdb.org). The molecular docking of two typical compounds with the highest inhibitory activities in benzyl imidazoles and benzyl carbazole synthesized (compounds **3** and **14**) was carried out using the CDocker protocol in accelrys discovery studio 2.1 software package (DS 2.1)[44]. Initially both the ligands and the enzyme (aromatase) were pretreated. The 3D structures of compounds **3** and **14** were generated with Cambridge Chem Bio Office 2008[45] and optimized with AM1 method. For enzyme preparation, the hydrogen atoms were added with the pH of the protein in the range of 6.5-8.5.

Model development: In order to give a systematic evaluation on benzyl imidazoles and benzyl carbazole as aromatase inhibitors and to explore more potent and selective aromatase inhibitors, a QSAR model was built using some 2D and 3D physiochemical properties as candidates of the descriptors which values were obtained from the calculation using the protocol of calculate molecular properties in DS 2.1. The compounds physiochemical properties include 2D (*AlogP, Molecular_Weight, Num_H_Acceptors, Num_H_Donors, Num_Rotatable Bonds, Molecular_Surface Area*, topological descriptors such as *BIC, SC_1, CIC, E_ADJ_equ, IAC_Total, IC and SIC, etc.*) and 3D (*Dipole, Jurs descriptors, shadow indices and Molecular_Volume, etc.*) parameters. pIC₅₀ (-lgIC₅₀ of the compounds was taken as the dependent variable. For the development of the QSAR model for the synthesized compounds **2-14**, the statistical techniques of genetic function algorithm and partial least squares (PLS) in Accelrys DS 2.1 were employed. In this study, all the 12 synthesized benzyl imidazoles derivatives and benzyl carbazole with definite IC₅₀ values were selected as the model dataset. More than 120

physiochemical properties such as *AlogP, Molecular_Weight, Dipole* and *Jurs* descriptors obtained from DS 2.1 as the initial independent descriptors to construct the QSAR model. For the system of a small quantity of samples with large parameters, the number of parameters should be reduced in order to make the obtained model with statistical meaning. In this study, genetic functional algorithm was employed to cut down the number of parameters and optimize the QSAR model. It solves problems in an analogous way of the organism's evolution. This method generates a series of potential solutions to a problem (the population of organisms) and then these solutions are modified and tested repeatedly until an approximate optimal solution is found[38,46]. During the optimizing process of genetic functional algorithm, Friedman lack-of-fit (LOF) was used as a scoring function[47] to control the model size to resist over-fitting, which is a problem often encountered in constructing statistical models. The smoothing factor of LOF was set to 0.5. For a given smoothing factor, the optimization of a QSAR model was considered to be obtained when descriptors usage in genetic functional algorithm became constant and independent of the increasing number of genetic functional algorithm crossover operations. All the descriptors in the QSAR trial descriptor pool were used as linear terms during

genetic functional algorithm to generate QSAR models. The obtained model was further optimized using G/PLS protocol in DS 2.5 and the correlation coefficient R^2 and the adjusted R^2(Adj-R^2), were taken as objective functions to select an equation. Leave-one-out cross-validation (R^2_{cv}) was employed to validate the predictivity of generated QSAR model equation.

RESULTS AND DISCUSSION

In order to examine the effect of substituents in the imidazoles on the aromatase ihibitory activity of benzyl imidazoles, imidazoles with substituents of 2-methyl, 2-ethyl, 2-ethyl-4-methyl, 4-carboxylate and benzo (carbazole) were synthesized. Considering the introduction of some substituents in the phenyl ring may improve the activity, infrequently used groups in aromatase inhibitors, carboxymethyl ester and its derivatives such as carboxymethyl, acetonitrile groups, carboxylation formamide were also considered, the structure of all the synthesized compounds are shown in **Schemes IV** and **V**.

The synthetic routes of compounds **1-9** are outlined in **Scheme-IV**. In order to check if imidazole was the optimal activity spacer for aromatase inhibition, benzyl imidazoles with different substitutional groups of compounds **10-13** were synthesized. The synthetic routes were illustrated in **Scheme-V**.
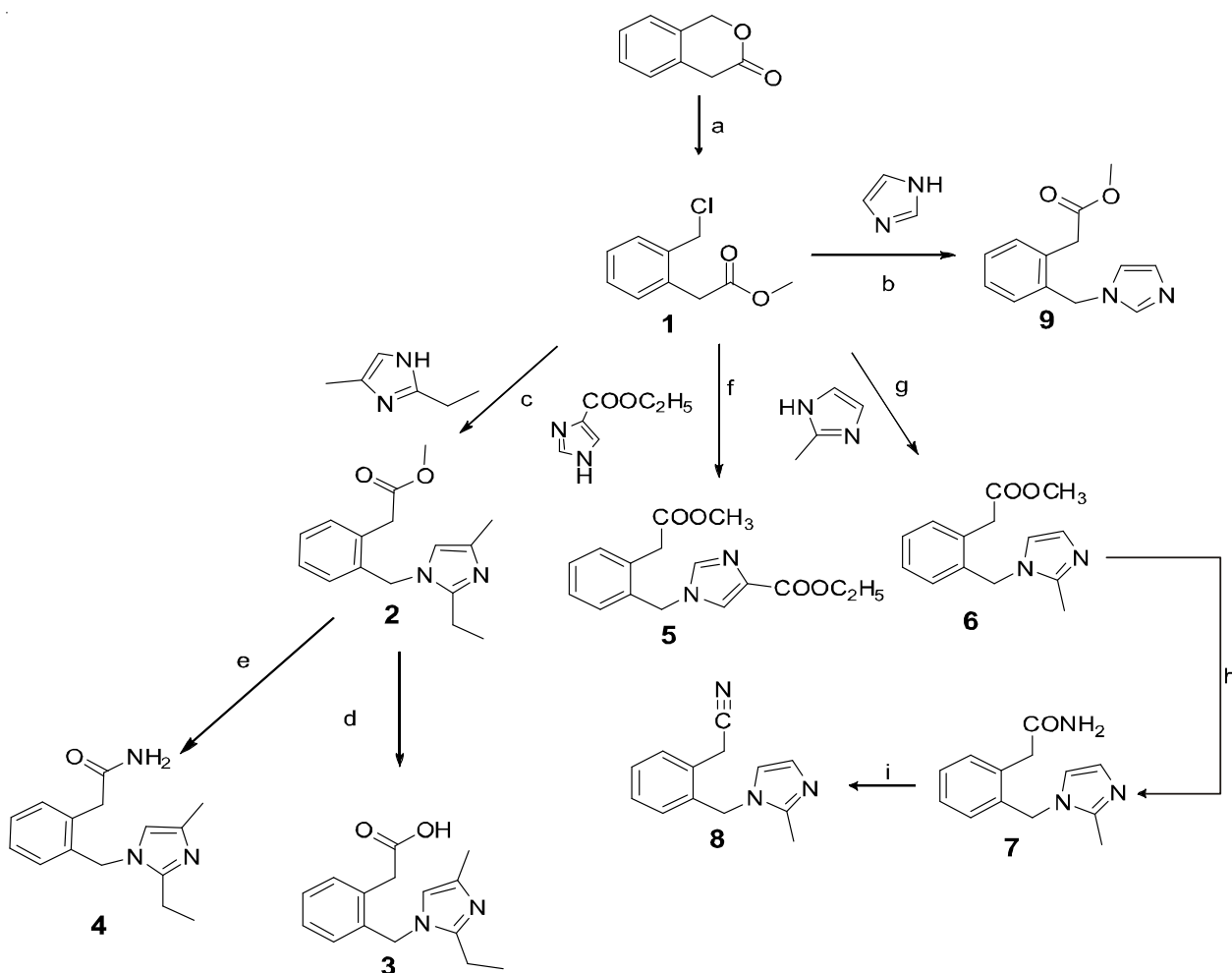
Scheme-IV: Synthesis of methyl 2-(2-(chloromethyl) phenyl)acetate and derivatives. Reagents and conditions: (a) SOCl₂, CH₃OH, -5 °C to 0 °C, 2h, 85.67 %; (b) K₂CO₃, CHCl₃, reflux, 10h, 73.22 %; (c) K₂CO₃, CHCl₃, reflux, 3h, 66.42 %; (d) 5 % NaOH, CH₃CH₂OH, reflux, 10h, 82.78 %; (e) NH₄OH, CH₃CH₂OH, rt, 72h, 63.51 %; (f) K₂CO₃, CHCl₃, reflux, 3h, 67.55 %; (g) K₂CO₃, CHCl₃, reflux, 5h, 75.80 %; (h) NH₄OH, CH₃CH₂OH, rt, 72h, 40.13 % and (i) POCl₃, K₂CO₃, EtOAc, reflux, 12h, 52.82 %
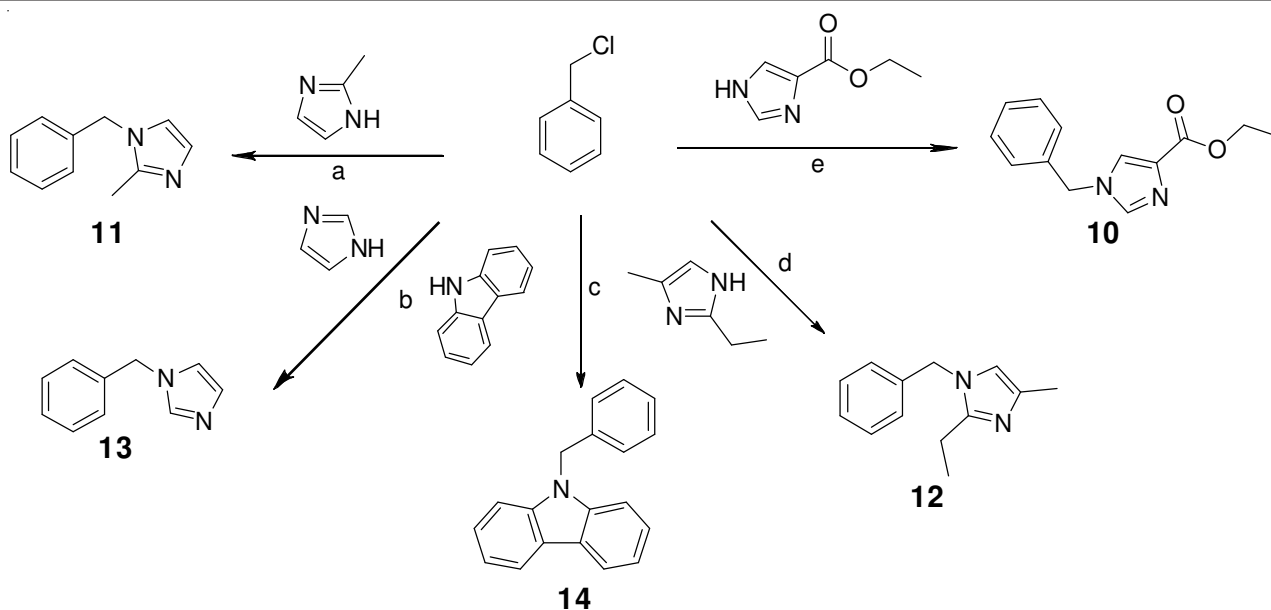
Scheme-V: Syntheses of benzylimidazole, benzylcarbazole and derivatives. Reagents and conditions: (a) K₂CO₃, CHCl₃, reflux, 6h, 75.56 %; (b) K₂CO₃, CHCl₃, reflux, 8h, 73.74 %; (c) K₂CO₃, DMF, rt to 40 °C, 8h, 56.41 %; (d) K₂CO₃, CHCl₃, reflux, 8h, 75.35 % and (e) K₂CO₃, CHCl₃, reflux, 5h, 79.16 %

Inhibitory activities: The in inhibitory activities of the target compounds were evaluated by a modification of the method of Luo *et al.*[41] and the results are listed in Table-1. It was found that if no substituent is existed in phenyl ring, the compounds have less even no inhibitory activity, such as compound **10** (IC$_{50}$ = 173.72 μM), **11** (IC$_{50}$ = 165.09 μM) and **13** (IC$_{50}$ = 140.96 μM). On the other hand, the introduction of some polar groups such as -CH$_2$CONH$_2$ (compound **4**) and -CH$_2$CN (compound **8**) also have less effect for the improvement of aromatase activity (IC$_{50}$ > 155.44 μM). However, the compounds with carboxyl and ester groups in phenyl ring show better inhibitory activity (the IC$_{50}$ values compounds **2** and **3** are 63.52 μM and 6.19 μM respectively). The introduction of alkyl groups in imidazole may improve the aromatase inhibitory activity. The inhibitory activity order is that compounds with 2-ethyl-4-methyl-1*H*-imidazole > compounds with 4-methyl-

1*H*-imidazole > compounds with 1*H*-imidazole. For example, compound **12** (IC$_{50}$ = 42.94 μM) > compound **11**(IC$_{50}$ = 165.09 μM) > compound **13** (IC$_{50}$ = 140.96 μM) and compound **6** (IC$_{50}$ = 6.96 μM) > compound **2** (IC$_{50}$ = 63.52 μM) > compound **9** (IC$_{50}$ = 173.72 μM). However, the introduction of ester in imidazole ring (compounds **5** and **6**) has little effect on the improvement of aromatase inhibitory activity.

Compared with the benzylimidazoles, the carbazole derivative (compound **14**) is more potent as aromatase inhibitors. Compound **14** displayed the most potent inhibitory activity among all the compounds, with an IC$_{50}$ of 2.72 μM.

Molecular docking and QSAR: The molecular docking of two typical compounds with the highest inhibitory activities in synthesized benzyl imidazoles and benzyl carbazole (compounds **3** and **14**) was carried out using the CDocker protocol in accelrys discovery studio 2.1 software package[42]. The poses

TABLE-1
OBSERVED AND PREDICTED AROMATASE INHIBITORY ACTIVITIES, PHYSIOCHEMICAL PROPERTIES OF
DIFFERENT STEROIDAL COMPOUNDS USED FOR THE CONSTRUCTION OF QSAR MODELS

Compounds	IC$_{50}$ (Obs[a]) (μM)	pIC$_{50}$ (Obs[a])	pIC$_{50}$ (Pred[b]) (M)	Residual	E_ADJ_equ	SC_1	Jurs_TASA	Shadow_XZ	CIC
2	63.53	4.197	4.207	-0.01	325.212	21	478.76	59.977	1
3	6.19	5.208	5.12	0.088	310.764	20	401.828	57.194	0.947
4	155.60	3.808	3.796	0.012	310.764	20	444.491	66.052	0.947
5	25.82	4.588	4.689	-0.101	354.413	23	395.107	67.836	0.818
6	6.97	5.157	5.159	-0.002	282.193	19	349.421	50.184	1.084
7	10.47	4.98	4.985	-0.005	268.078	18	355.493	47	1.265
8	189.23	3.723	3.818	-0.095	240.215	17	397.365	49.001	1.219
9	173.78	3.76	3.677	0.083	254.084	18	388.828	52.13	1.176
10	173.78	3.76	3.679	0.081	254.084	18	384.577	60.1	0.868
11	194.98	3.71	3.77	-0.06	186.117	14	353.371	47.043	1.231
12	42.95	4.367	4.4	-0.033	226.477	16	429.07	54.279	0.717
13	140.93	3.851	3.824	0.027	160	13	332.011	43.889	1.801
14	2.72	5.565	5.551	0.014	384	23	482.628	57.057	1.683
Aminoglutethimide	33.11	4.480							
Tamoxifen	3.24	5.490							

[a] Obs, observed, [b] Pred, predicted

of compounds **3** and **14** within the important amino acid residues of human placental aromatase are shown in Figs. 2 and 3 (the molecule in pink stick is haem and that in gray stick is compound **3**. The green dash line refers to hydrogen bond). The results showed that the oxygens in carbonyl of compound **3** coordinated with the Fe(III) in haem molecule. The distances of O-Fe are 2.304 Å and 2.645 Å respectively. There are also hydrogen bonds formed between the hydroxyl group in compound **3** and carbonyl group in ALA306, hydroxyl group in THR310. However, there are no hydrogen bond formed between compound **14** and CYP19. It may be the hydrophobic interaction plays the main role for their interaction. The highest bioactivity of compound **14** may be attributed to the strong hydrophobic interaction between them.
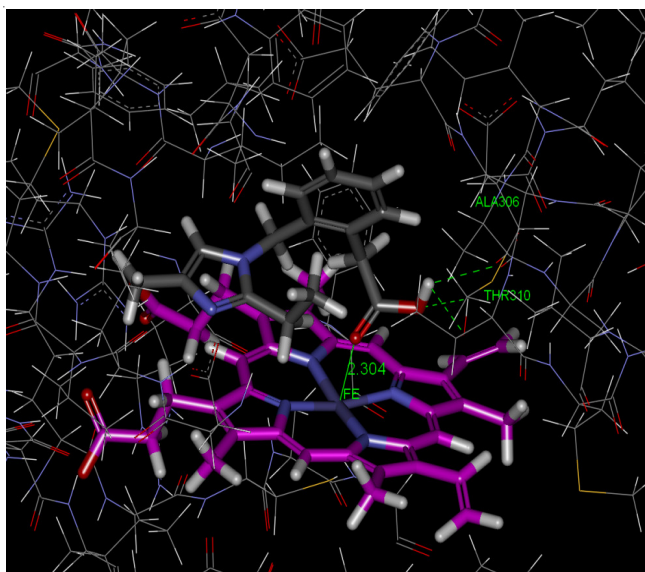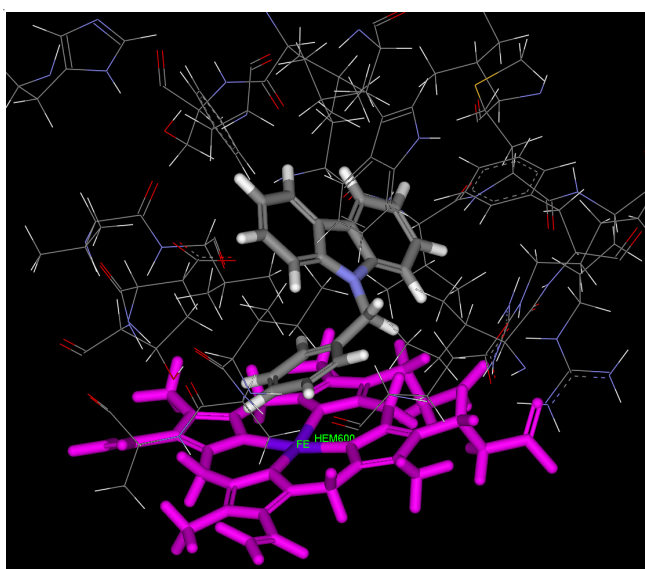

Fig. 2

Fig. 3

The QSAR model was optimized and descriptors were reduced with the genetic function approximation. Some initial parameters for genetic function approximation are listed in Table-2.

TABLE-2
INITIAL PARAMETERS FOR GFA OPTIMIZATION

Population	Maximum generations	Scoring function	Scaled LOF smoothness parameter	Mutation probability
100	5000	Friedman LOF	0.50	0.100

Under this condition, the model with the Friedman LOF value was acquired as follows:

$$pIC_{50} = 17.2938 + 0.124234 \times E\_ADJ\_equ - 1.59464$$
$$\times SC\_1 - 0.00726719 \times Jurs\_TASA - 0.114515$$
$$\times Shadow\_XZ + 2.86407 \times < 1.23168 - CIC > \quad (1)$$

The sample number is 13 Friedman LOF = 0.00359, R^2 = 0.9914, Adj-R^2 = 0.9853, R^2cv = 0.9639, F = 161.8. The observed and predicted pIC_{50} results and the values of physiochemical properties of the 13 ligands for Eqn. (1) are listed in Table-1. In our study, R^2, Adj-R^2, R^2cv and F-value were used to evaluate the regression model. Eqn. (1) can explain 98.33 % of the variance (Adj-R^2) while it could predict 97.06 % of the variance (R^2cv). F > F (a = 0.05) = 4 shows that the model is in the confidence interval of 95 %. The difference between R^2 and Pred-R^2 values is not very high (less than 0.3)[34]. It can be seen from Eq. (1) that *SC_1, Jurs_TASA, Shadow_XZ* and *CIC* have negative contribution to the bioactivity of the ligands, however, only *E_ADJ_equ*, has the positive effect on the bioactivity of the ligands. The plot of the observed pIC_{50} *vs.* the predicted data for compounds **2-14** is shown in Fig. 4. The high correlation coefficient (R^2 = 0.9914) for the regression of predicted versus measured pIC_{50} for compounds **2-14** and all the points locating near the y = x axis (the solid line) indicated that the predicted data by this model is in accordance well with the experimental results.
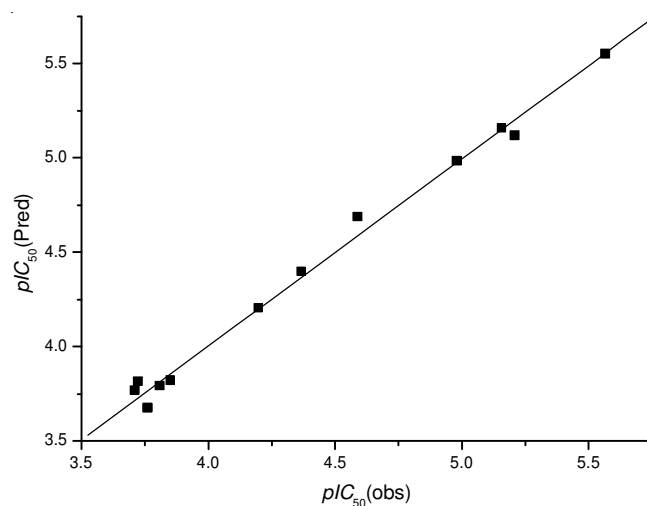

Fig. 4

The standardized regression coefficient for each variable is 11.84322, 7.249866, 0.520296, 0.031171 and 0.805255 respectively. Therefore, the relative importance of the descriptors according to their standardized regression coefficients is in the following order:

$$E\_ADJ\_equ > SC\_1 >> <1.23168\text{-}CIC>>$$
$$Jurs\_TASA > Shadow\_XZ$$

It was found that *E_ADJ_equ* and *SC_1* play the key role for the bioactivity of the ligands.

Conclusion

Benzyl carbazole, 12 benzyl imidazole derivatives with different substituents on both phenyl, imidazole rings were synthesized and their aromatase inhibitory were evaluated with fluorescent substrate detection method. The results showed that the compounds with carboxyl and ester groups in phenyl ring show better inhibitory activity. The introduction of alkyl groups in imidazole may improve the aromatase inhibitory activity. The obtained QSAR model with genetic functional algorithm gives comparatively better correlation coefficient and cross-validation correlation coefficient.

REFERENCES

1. R.W. Brueggemeier, J.C. Hackett and E.S. Diaz-Cruz, *Endocr. Rev.*, **26**, 331 (2005).
2. C. DeSantis, R. Siegel, P. Bandi and A. Jemal, *CA Cancer J. Clin.*, **61**, 408 (2011).
3. H. Yu, R. Harris, Y. -T. Gao, R. Gao and E. Wynder, *Int. J. Epidemiol.*, **20**, 76 (1991).
4. L. Yang, L.D. Li, Y.D. Chen and D.M. Parkin, *Zhonghua Zhong Liu Za Zhi*, **28**, 438 (2006).
5. R.J. Santen, *Steroids*, **50**, 575 (1987).
6. M.S. Mahendroo, C.R. Mendelson and E.R. Simpson, *J. Biol. Chem.*, **268**, 19463 (1993).
7. E.R. Simpson, M.S. Mahendroo, G.D. Means, M.W. Kilgore, M.M. Hinshelwood, S. Graham-Lorence, B. Amarneh, Y. Ito, C.R. Fisher and M.D. Michael, *Endocr. Rev.*, 15, 342 (1994).
8. H. Lefebvre, V. Contesse, C. Delarue, C. Soubrane, A. Legrand, J.M. Kuhn, L.M. Wolf and H. Vaudry, *J. Clin. Endocrinol. Metab.*, **77**, 1662 (1993).
9. W.R. Miller and J. O'Neill, *Steroids*, **50**, 537 (1987).
10. J. McNulty, J. Nair, N. Vurgun, B.R. DiFrancesco, C.E. Brown, B. Tsoi, D.J. Crankshaw and A.C. Holloway, *Bioorg. Med. Chem. Lett.*, **22**, 718 (2012).
11. T.R. Evans, E.D. Salle, G. Ornati, M. Lassus, M.S. Benedetti, E. Pianezzola and R.C. Coombes, *Cancer Res.*, **52**, 5933 (1992).
12. D.P. McDonnell and S.E. Wardell, *Curr. Opin. Pharmacol.*, **10**, 620 (2010).
13. A. Buzdar and A. Howell, *Clin. Cancer Res.*, **7**, 2620 (2001).
14. P.E. Goss and K. Strasser, *J. Clin. Oncol.*, **19**, 881 (2001).
15. A. Hamilton and M. Volm, Oncology (Williston Park), 15, 965-972; discussion 972, 977-969 (2001).
16. R. Murray, *Cancer Chemother. Pharmacol.*, **48**, 259 (2001).
17. J.N. Ingle, *Clin. Cancer Res.*, **7**, 4392s (2001).
18. S. Gobbi, C. Zimmer, F. Belluti, A. Rampa, R.W. Hartmann, M. Recanatini and A. Bisi, *J. Med. Chem.*, **53**, 5347 (2010).
19. L.L. Woo, T. Jackson, A. Putey, G. Cozier, P. Leonard, K.R. Acharya, S.K. Chander, A. Purohit, M.J. Reed and B.V.L. Potter, *J. Med. Chem.*, **53**, 2155 (2010).
20. A. Stefanachi, A.D. Favia, O. Nicolotti, F. Leonetti, L. Pisani, M. Catto, C. Zimmer, R.W. Hartmann and A. Carotti, *J. Med. Chem.*, **54**, 1613 (2011).
21. P.M. Wood, L. Woo, M.P. Thomas, M.F. Mahon, A. Purohit and B.V.L. Potter, *Chem. Med. Chem.*, **6**, 1423 (2011).
22. K. Bonfield, E. Amato, T. Bankemper, H. Agard, J. Steller, J.M. Keeler, D. Roy, A. McCallum, S. Paula and L. Ma, *Bioorg. Med. Chem.*, **20**, 2603 (2012).
23. F. Caporuscio, G. Rastelli, C. Imbriano and A. Del Rio, *J. Med. Chem.*, **54**, 4006 (2011).
24. D. Ghosh, J. Lo, D. Morton, D. Valette, J. Xi, J. Griswold, S. Hubbell, C. Egbuta, W. Jiang, J. An and H.M.L. Davies, *J. Med. Chem.*, **55**, 8464 (2012).
25. T. Liu, Y. Lin, X. Wen, R.N. Jorissen and M.K. Gilson, Nucleic Acids Res., 35(Database), D198 (2007).
26. L. Roumen, J.W. Peeters, J.M. Emmen, I.P.E. Beugels, E.M.G. Custers, M. de Gooyer, R. Plate, K. Pieterse, P.A.J. Hilbers, J.F.M. Smits, J.A.J. Vekemans, D. Leysen, H.C.J. Ottenheijm, H.M. Janssen and J.J.R. Hermans, *J. Med. Chem.*, **53**, 1712 (2010).
27. J. Doiron, A.H. Soultan, R. Richard, M.M. Touré, N. Picot, R. Richard, M. Cuperlovic-Culf, G.A. Robichaud and M. Touaibia, *Eur. J. Med. Chem.*, **46**, 4010 (2011).
28. C.P. Owen, S. Dhanani, C.H. Patel, I. Shahid and S. Ahmed, *Bioorg. Med. Chem. Lett.*, **16**, 4011 (2006).
29. P.P. Roy and K. Roy, *J. Mol. Model.*, **16**, 1597 (2010).
30. B. Balogh, B. Jójárt, Z. Wágner, P. Kovacs, G. Mate, K. Gyires, Z. Zadori, G. Falkay, A. Marki and B. Viskolcz, *Neurochem. Int.*, **51**, 268 (2007).
31. Y.C. Martin, Quantitative Drug Design: A Critical Introduction, M. Dekker, New York (1978).
32. R.R. Pissurlenkar, M.S. Shaikh and E.C. Coutinho, *J. Mol. Model.*, **13**, 1047 (2007).
33. P. Zhou and Z. Li, *Sci. China B*, **50**, 568 (2007).
34. A. Liu, H. Guang, L. Zhu, G.H. Du, S.M.Y. Lee and Y.T. Wang, *Sci. China C Life Sci.*, **50**, 726 (2007).
35. J.T. Leonard and K. Roy, *Bioorg. Med. Chem.*, **14**, 1039 (2006).
36. X. Wang, Q. Xiao and S. Cui, *Sci. China B*, **48**, 156 (2005).
37. I.L. Ruiz, M. Urbano-Cuadrado and M.Á. Gómez-Nieto, *J. Math. Chem.*, **43**, 1549 (2008).
38. Y. Dai, Q. Wang, X. Zhang, S. Jia, H. Zheng, D. Feng and P. Yu, *Eur. J. Med. Chem.*, **45**, 5612 (2010).
39. D. Ghosh, J. Griswold, M. Erman and W. Pangborn, *Nature*, **457**, 219 (2009).
40. G.F. Yang, H.T. Lu, Y. Xiong and C.-G. Zhan, *Bioorg. Med. Chem.*, **14**, 1462 (2006).
41. L. Luo, L. Shen, F. Sun, Y. Dai, H. Zheng, Z. Ma, Y. Xu and Z. Guo, *Anal. Methods*, **4**, 230 (2012).
42. Discovery Studio. Accelrys Software Inc, San Diego, CA.
43. D.J. Ritchie, US Patent 6048998(2000).
44. C.P. Miller, P. Bhaket, N. Muthukaman, C.R. Lyttle, M. Shomali, K. Gallacher, C. Slocum and G. Hattersley, *Bioorg. Med. Chem. Lett.*, **20**, 7516 (2010).
45. Chem. Bio. Office Ultra Version 11.0, Cambridge Soft. Inc., Cambridge, USA.
46. Y. Dai, N. Chen, Q. Wang, H. Zheng, X. Zhang, S. Jia, L. Dong and D. Feng, *Iran. J. Pharm. Res.*, **11**, 807 (2012).
47. J.H. Friedman, *Ann. Statistics*, **19**, 1 (1991).