EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

## Original article:

# PROBING THE ORIGINS OF AROMATASE INHIBITORY ACTIVITY OF DISUBSTITUTED COUMARINS VIA QSAR AND MOLECULAR DOCKING

Apilak Worachartcheewan[1], Naravut Suvannang[1], Supaluk Prachayasittikul[1], Virapong Prachayasittikul[2], Chanin Nantasenamat[1,2,*]

[1] Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[2] Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

* Corresponding author: E-mail: chanin.nan@mahidol.ac.th (C.N.); Phone: +66 2 441 4371; Fax: +66 2 441 4380

## ABSTRACT

This study investigated the quantitative structure-activity relationship (QSAR) of imidazole derivatives of 4,7-disubstituted coumarins as inhibitors of aromatase, a potential therapeutic protein target for the treatment of breast cancer. Herein, a series of 3,7- and 4,7-disubstituted coumarin derivatives (1-34) with $R_1$ and $R_2$ substituents bearing aromatase inhibitory activity were modeled as a function of molecular and quantum chemical descriptors derived from low-energy conformer geometrically optimized at B3LYP/6-31G(d) level of theory. Insights on origins of aromatase inhibitory activity was afforded by the computed set of 7 descriptors comprising of F10[N-O], Inflammat-50, Psychotic-80, H-047, BELe1, B10[C-O] and MAXDP. Such significant descriptors were used for QSAR model construction and results indicated that model 4 afforded the best statistical performance. Good predictive performance were achieved as verified from the internal (comprising the training and the leave-one-out cross-validation (LOO-CV) sets) and external sets affording the following statistical parameters: $R^2_{Tr}$ = 0.9576 and $RMSE_{Tr}$ = 0.0958 for the training set; $Q^2_{CV}$ = 0.9239 and $RMSE_{CV}$ = 0.1304 for the LOO-CV set as well as $Q^2_{Ext}$ = 0.7268 and $RMSE_{Ext}$ = 0.2927 for the external set. Significant descriptors showed correlation with functional substituents, particularly, $R_1$ in governing high potency as aromatase inhibitor. Molecular docking calculations suggest that key residues interacting with the coumarins were predominantly lipophilic or non-polar while a few were polar and positively-charged. Findings illuminated herein serve as the impetus that can be used to rationally guide the design of new aromatase inhibitors.

**Keywords:** Coumarin, aromatase, aromatase inhibitor, QSAR, data mining, molecular docking

## INTRODUCTION

Breast cancer is commonly found in women and is reported as the second leading cause of women death (Desantis et al., 2011). Estrogen is associated with the development of breast cancer by activating intracellular signaling cascades (Yager and Davidson, 2006). Aromatase is an enzyme involved in the catalysis of androgens (i.e. androstenedione) to estrogens (i.e. estradiol). Therefore, the ability to inhibit such enzyme

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

offers great therapeutic benefits for the treatment of breast cancer (Narashimamurthy et al., 2004). Aromatase (i.e. 19A1) is a member of the cytochrome P450 family and is comprised of 503 amino acids. The proximal ligand of aromatase (i.e. Cys437) is coordinated to the heme prosthetic group and it was previously reported that the presence of the characteristic cysteine proximal ligand is crucial for the catalytic activity observed in cytochrome P450s (Auclair et al., 2001; Yoshioka et al., 2001). Aromatase inhibitors comprising of steroidal and non-steroidal compounds have either been synthesized by introducing or modifying functional groups on the core structure of prototype lead compounds (Ferlin et al., 2013; Nativelle-Serpentini et al., 2004; Neves et al., 2009; Stefanachi et al., 2011; Varela et al., 2012) or extracted from natural sources (Balunas et al., 2008). In addition, computational analysis such as quantitative structure-activity relationship (QSAR), molecular docking and modeling had previously been employed for constructing models and interpreting the interaction between compounds of interest with the aromatase enzyme (Bheemanapalli et al., 2013; Galeazzi and Massaccesi 2012; Ghosh et al., 2012; Nantasenamat et al., 2013a, b; Narayana et al., 2012).

Coumarin is a naturally occurring bioactive benzopyrone found in many plant species (Venugopala et al., 2013). It represents an important structural scaffold in medicinal chemistry as it can afford a wide range of bioactivities such as anti-bacterial (Nagamallu and Kariyappa, 2013), anti-cancer (Wu et al., 2014), anti-inflammatory (Hemshekhar et al., 2013), anti-oxidant (Guinez et al., 2013) as well as neuroprotective (Sun et al., 2013) properties. Furthermore, 7-hydroxycoumarin is the primary metabolite of coumarin in human. A number of 7-oxy substituted coumarin analogs have been shown to exhibit an array of bioactivities (Venugopala et al., 2013). Recently, 3,7- and 4,7-disubstituted coumarin derivatives (**1-34**, Figure 1) have been reported to display aromatase inhibitory potency in the nanomolar range (Stefanachi et al., 2011).

The aim of this study is to explain the origins of aromatase inhibitory activity via QSAR modeling and molecular docking. QSAR models were built as a function of significant descriptors accounting for aromatase inhibitory activity. Insights into structure-activity relationship are also discussed and it is anticipated that such information could serve as a pertinent guideline for the rational design of novel aromatase inhibitors based on the coumarin chemotype.

## MATERIALS AND METHODS

### *Data set*

A set of coumarin derivatives with reported aromatase inhibitory activity was taken from the work of Stefanachi et al. (2011). Briefly, the aromatase inhibitory activity was determined *in vitro* using human placental microsomes as the source of aromatase while [1$\beta$-$^3H$] androstenedione was used as the substrate. $IC_{50}$ values were logarithmically transformed to $pIC_{50}$ as summarized by the following equation:

$$pIC_{50} = -\log_{10}(IC_{50}) \quad [1]$$

Chemical structures are depicted in Figure 1 while its corresponding descriptors and bioactivity are shown in Table 1. Schematic workflow of this study is displayed in Figure 2.

### *Geometrical optimization and descriptors calculation*

Molecular structures were drawn using ChemAxon MarvinSketch version 6.2.1 (ChemAxon Ltd., 2014) and converted to the appropriate file format with Babel version 3.3.0 (OpenEye Scientific Software, 2014). An initial geometrical optimization was performed at the semi-empirical Austin Model 1 (AM1) level followed by further refinement at the density functional theory (DFT) level using the Becke's three parameter hybrid method and the Lee-Yang-Parr (B3LYP) functional together with the 6-31g(d) basis set using the Gaussian 09 software package

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

(Frisch et al., 2009). A set of 6 quantum chemical descriptors were obtained from the aforementioned low-energy conformer that included the following: dipole moment ($\mu$), the total energy, the highest occupied molecular orbital energy (HOMO), the lowest unoccupied molecular orbital (LUMO), difference in energy values of HOMO and LUMO (HOMO-LUMO) and the mean absolute atomic charge ($Q_m$) (Karelson et al., 1996; Thanikaivelan et al., 2000). Furthermore, the low-energy conformer were also subjected to the generation of additional descriptors from the Dragon software package, version 5.5 (Talete, 2007) to derive a set of 3,224 molecular descriptors comprising 22 categories: 48 Constitutional descriptors, 119 Topological descriptors, 47 Walk and path counts, 33 Connectivity indices, 47 Information indices, 96 2D autocorrelation, 107 Edge adjacency indices, 64 Burden eigenvalues, 21 Topological charge indices, 44 Eigenvalue-based indices, 41 Randic molecular profiles, 74 Geometrical descriptors, 150 RDF descriptors, 160 3D-MoRSE descriptors, 99 WHIM descriptors, 197 GETAWAY descriptors, 154 Functional group counts, 120 Atom-centered fragments, 14 Charge descriptors, 29 Molecular properties, 780 2D binary fingerprints and 780 2D frequency fingerprints.

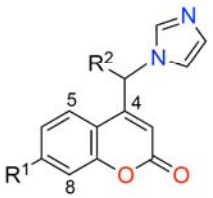
### *Descriptors selection*

Constant variables from the initial set of 3,224 molecular descriptors obtained from the Dragon software were subjected to removal and subsequently combined with a set of 6 quantum chemical descriptors. Significant descriptors correlated with the aromatase inhibitory activity were derived from stepwise multiple linear regression (Worachartcheewan et al., 2014) using SPSS Statistics version 18.0 (SPSS Inc., USA). Intercorrelation matrix of descriptors was constructed from Pearson's correlation coefficient values as to deduce the presence of variable redundancy.

### *Generation of internal and external sets*

The data set was divided into 2 major subsets in which an external set was obtained by randomly selecting 15 % of the data samples from the full data set while the remaining 85 % served as the internal set (Nantasenamat et al., 2013a). The internal set was used to internally assess the predictive performance of QSAR models by using

**1** $R^1 = OCH_2C_6H_5$, $R^2 = H$
**2** $R^1 = OCH_2C_6H_4$-$CH_3$-*m*, $R^2 = H$
**3** $R^1 = OCH_2C_6H_4$-F-*m*, $R^2 = H$
**4** $R^1 = OCH_2C_6H_4$-Cl-*m*, $R^2 = H$
**5** $R^1 = OCH_2C_6H_4$-$CF_3$-*m*, $R^2 = H$
**6** $R^1 = OCH_2C_6H_4$-$OCF_3$ -*m*, $R^2 = H$
**7** $R^1 = OCH_2C_6H_4$- $NO_2$-*m*, $R^2 = H$
**8** $R^1 = OCH_2C_6H_4$-F-*p*, $R^2 = H$
**9** $R^1 = OCH_2C_6H_4$-Cl-*p*, $R^2 = H$
**10** $R^1 = OCH_2C_6H_4$-$OCH_3$-*p*, $R^2 = H$
**11** $R^1 = OCH_2C_6H_4$-$OCF_3$-*p*, $R^2 = H$
**12** $R^1 = OCH_2C_6H_3$-3,5 di F, $R^2 = H$
**13** $R^1 = OCH_2C_6H_3$-3,4 di F, $R^2 = H$
**14** $R^1 = OC_6H_5$, $R^2 = H$
**15** $R^1 = OC_6H_4$-F-*m*, $R^2 = H$
**16** $R^1 = OC_6H_4$-Cl-*m*, $R^2 = H$
**17** $R^1 = OC_6H_4$-$OCH_3$-*m*, $R^2 = H$
**18** $R^1 = OC_6H_4$-$CH_3$-*p*, $R^2 = H$
**19** $R^1 = OC_6H_4$-Cl -*p*, $R^2 = H$
**20** $R^1 = OC_6H_4$- CN-*p*, $R^2 = H$
**21** $R^1 = OC_6H_4$-$COCH_3$-*p*, $R^2 = H$
**22** $R^1 = OC_6H_4$-$N(CH_3)_2$-*p*, $R^2 = H$




**1-34**

**23** $R^1 = OC_6H_3$-3,5 di F, $R^2 = H$
**24** $R^1 = OC_6H_3$-3,4 di F, $R^2 = H$
**25** $R^1 = NHC_6H_5$, $R^2 = H$
**26** $R^1 = OH$, $R^2 = H$
**27** $R^1 = OCH_3$, $R^2 = H$
**28** $R^1 = OCH_3$, $R^2 = C_6H_5$
**29** $R^1 = OC_6H_5$, $R^2 = C_6H_5$
**30** $R^1 = OC_6H_3$-3,4 di F, $R^2 = C_6H_5$
**31** $R^1 = OC_6H_5$, $R^2 = C_6H_4$-Cl-*p*
**32** $R^1 = OC_6H_5$, $R^2 = C_6H_4$-CN-*p*
**33** $R^1 = OCH_3$, $R^2 = H$
**34** $R^1 = OC_6H_5$, $R^2 = C_6H_5$

**Figure 1:** Chemical structures of disubstituted coumarins **1**-**34**

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

**Table 1:** Molecular descriptors and aromatase inhibitory activity of coumarins **(1-34)**

| Cpd. | Descriptors | | | | | | | Activity | |
|---|---|---|---|---|---|---|---|---|---|
| | F10[N-O] | Inflammat-50 | Psychotic-80 | H-047 | BELe1 | B10 [C-O] | MAXDP | $IC_{50}$ (μM) | $pIC_{50}$ |
| **1** | 0 | 0 | 1 | 13 | 1.920 | 1 | 4.891 | 0.150 | 6.824 |
| **2** | 0 | 0 | 1 | 12 | 1.922 | 1 | 4.929 | 0.114 | 6.943 |
| **3** | 0 | 0 | 1 | 12 | 1.920 | 1 | 5.234 | 0.113 | 6.947 |
| **4** | 0 | 0 | 1 | 12 | 1.920 | 1 | 4.914 | 0.130 | 6.886 |
| **5** | 0 | 0 | 1 | 12 | 1.921 | 1 | 4.916 | 0.235 | 6.629 |
| **6** | 0 | 0 | 0 | 12 | 1.920 | 1 | 4.937 | 0.207 | 6.684 |
| **7** | 0 | 0 | 1 | 12 | 1.922 | 1 | 4.929 | 0.141 | 6.851 |
| **8** | 0 | 0 | 1 | 12 | 1.920 | 1 | 4.947 | 0.267 | 6.573 |
| **9** | 0 | 0 | 1 | 12 | 1.920 | 1 | 4.910 | 0.178 | 6.750 |
| **10** | 0 | 0 | 1 | 15 | 1.920 | 1 | 4.940 | 0.127 | 6.896 |
| **11** | 0 | 0 | 0 | 12 | 1.920 | 1 | 4.930 | 0.481 | 6.318 |
| **12** | 0 | 1 | 0 | 11 | 1.923 | 1 | 4.870 | 0.169 | 6.772 |
| **13** | 0 | 0 | 1 | 11 | 1.920 | 1 | 5.271 | 0.165 | 6.783 |
| **14** | 0 | 1 | 0 | 11 | 1.923 | 1 | 4.870 | 0.051 | 7.292 |
| **15** | 0 | 1 | 0 | 10 | 1.923 | 1 | 5.262 | 0.072 | 7.143 |
| **16** | 0 | 1 | 0 | 10 | 1.923 | 1 | 4.897 | 0.072 | 7.143 |
| **17** | 0 | 0 | 1 | 13 | 1.924 | 1 | 4.937 | 0.292 | 6.535 |
| **18** | 0 | 0 | 1 | 10 | 1.926 | 1 | 4.907 | 0.690 | 6.161 |
| **19** | 0 | 1 | 0 | 10 | 1.923 | 1 | 4.893 | 0.112 | 6.951 |
| **20** | 1 | 1 | 0 | 10 | 1.925 | 1 | 4.910 | 0.164 | 6.785 |
| **21** | 0 | 0 | 1 | 10 | 1.926 | 1 | 4.939 | 0.296 | 6.529 |
| **22** | 0 | 0 | 1 | 16 | 1.925 | 1 | 4.967 | 0.081 | 7.092 |
| **23** | 0 | 1 | 0 | 9 | 1.923 | 1 | 5.283 | 0.070 | 7.155 |
| **24** | 0 | 1 | 0 | 9 | 1.923 | 1 | 5.299 | 0.047 | 7.328 |
| **25** | 0 | 1 | 0 | 11 | 1.935 | 1 | 4.890 | 0.105 | 6.979 |
| **26** | 0 | 0 | 0 | 6 | 1.914 | 0 | 4.471 | 3.750 | 5.426 |
| **27** | 0 | 0 | 0 | 9 | 1.916 | 0 | 4.600 | 0.280 | 6.553 |
| **28** | 0 | 0 | 1 | 13 | 1.955 | 1 | 5.177 | 0.455 | 6.342 |
| **29** | 0 | 0 | 1 | 15 | 1.958 | 1 | 5.447 | 0.067 | 7.174 |
| **30** | 0 | 0 | 1 | 13 | 1.957 | 1 | 5.503 | 0.317 | 6.499 |
| **31** | 0 | 0 | 0 | 14 | 1.957 | 1 | 5.480 | 0.532 | 6.274 |
| **32** | 2 | 0 | 1 | 14 | 1.962 | 1 | 5.504 | 4.010 | 5.397 |
| **33** | 1 | 0 | 0 | 9 | 1.908 | 1 | 4.918 | 2.820 | 5.550 |
| **34** | 1 | 0 | 1 | 15 | 1.945 | 1 | 5.972 | 0.313 | 6.504 |

it as training and leave-one-out cross-validated (LOO-CV) sets. LOO-CV refers to the leaving out of one data sample as the testing set while using the remaining *N*-1 samples as the training set. The process was iteratively repeated until all samples were used as the testing set. As the name implies, the external set was used to externally assess the predictive performance of the QSAR model.

### *Multiple linear regression*

Multiple linear regression (MLR) ws employed for constructing QSAR models as conceptually summarized by the following equation:

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

$$Y = m_1x_1 + m_2x_2 + \cdots + m_nx_n + b \quad [2]$$

where $Y$ is the aromatase inhibitory activity ($pIC_{50}$), $m$ is the regression coefficient value of descriptors, $x$ is the descriptor and $b$ repreents the $y$-intercept value. MLR calculations were used as implemented by the Waikato Environment for Knowledge Analysis (WEKA) software, version 3.4.5 (Witten et al., 2011).

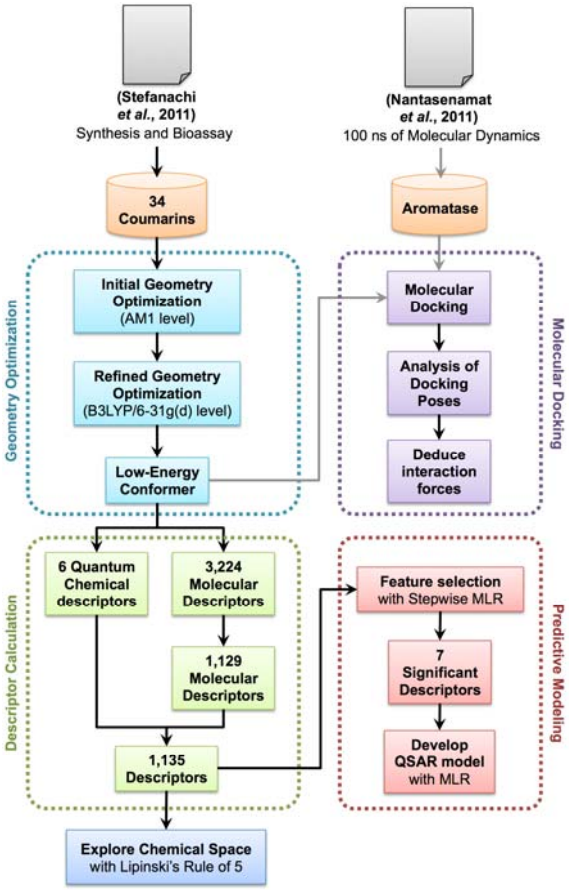



**Figure 2:** Schematic workflow of QSAR and molecular docking studies performed herein

### *Statistical analysis*

Evaluation of the predictive performance of QSAR models was performed using squared correlation coefficient for the training ($R^2_{Tr}$) and cross-validated sets ($Q^2_{CV}$) as well as the root mean squared error for the training ($RMSE_{Tr}$) and cross-validated sets ($RMSE_{CV}$). Furthermore, the Fisher (F) ratio as well as the difference of $R^2$ and $Q^2$ ($R^2$–$Q^2$) was used to assess the predictive quality of constructed models. Moreover, squared correlation coefficient (Q2Ext) and root mean squared error (RMSEExt) of external set were used as independent validation of the constructed QSAR model (Nantasenamat et al., 2013a).

### *Molecular docking*

The set of 34 coumarin compounds was used as ligands for docking to the X-ray crystallographic structure of aromatase (PDB id 3EQM). Prior to docking, both protein and ligand structures were subjected to a series of pre-processing. A snapshot of the aromatase structure was obtained from 100 ns of molecular dynamics simulation using the AMBER03 force field under YASARA 10.11.28 (Krieger et al., 2002, 2003) as reported in our previous investigation (Suvannang et al., 2011). Subsequently, atomic charges of the heme prosthetic group were parameterized according to density functional calculations reported by Favia et al. (2006). Low-energy conformer of ligands as obtained from previously mentioned geometrical optimization. Kollman and Gasteiger charges were finally assigned to the protein and ligands, respectively, using AutoDockTools (Morris et al., 2009).

Docking calculatons were performed using AutoDock version 4.2 (Morris et al., 2009). A grid box was constructed using AutoGrid and centered at the binding cavity using x,y,z coordinates of 84.0236, 49.6568, 50.0293 where it is encapsulated by a box size of 52 × 66 × 82 Å points with spacing of 0.3759 Å. Each docking pose was obtained from 100 independent runs based on 150 randomly placed individuals in the population. The docking simulation made use of the Lamarckian Genetic Algorithm (Morris et al., 1998) in searching for low-energy binding orientation. Translation step of 2.0 Å, mutation rate of 0.02, crossover rate of 0.8, local search rate of 0.06 and maximum energy evaluation of 250,000,000 were utilized in the present docking simulation. Docked conformations were clustered using an RMSD tolerance of 2.0 Å. Re-docking of the

co-crystallized substrate was performed as to evaluate the validity of the docking protocol.

Post-docking analyses were carried out using AutoDockTools and graphical images were prepared using PyMOL version 0.99 (Delano, 2002).

## RESULTS AND DISCUSSION

### *Chemical space of investigated coumarins*

Stefanachi et al. (2011) reported the synthesis and determination of aromatase inhibitory activity for a set of coumarin derivatives (**1**-**34**) as shown in Figure 1. The coumarin core structure featured various substituents (i.e. halogen, methoxy, aryloxy and imidazole) that conferred its bioactivity against aromatase (Table 1). In gaining further insights on the general features of potency afforded by these coumarins, compounds having $pIC_{50}$ values greater than 7 (< 0.081 nM) were considered active, less than 6 as inactive (> 2.82 μM) and values in between were considered to have intermediate activity. It can be seen that the potency of this set of compounds was relatively robust as deduced from $IC_{50}$ values in the range of 47 nM to 690 nM for 31 out of the 34 compounds. The remaining three compounds afforded poor $IC_{50}$ values in the range of 2.820 and 4.010 μM.

Applying the Lipinski's rule-of-five on this set of compounds revealed its compliance in which there were less than 5 H-bond donors, less than 10 H-bond acceptors, molecular weight (MW) of less than 500 Da and Ghose–Crippen octanol–water partition coefficient (ALogP) of less than 5 (Supplementary Information). An exploratory data analysis of the three aforementioned sub-classes revealed that intermediate and active subclasses were generally larger than their inactive counterpart with MW of 361.246 ± 39.971, 353.146 ± 23.348 and 305.997 ± 98.512 Da, respectively. However, it should be noted that of the three compounds in the inactive sub-class, **32** had significantly higher MW of 419.46 Da than the remaining two with MW of 242.25 and 256.28 Da for **26** and **33**, respectively. The number of H-bond donors does not appear to be crucial for its bioactivity as can be seen from the sparseness of their occurrence in the coumarins in which a total of two compounds had one H-bond donor belonging to the inactive (**26**) and intermediate (**25**) sub-class. As for the number of H-bond acceptors, the intermediate and active sub-classes had slightly higher mean values of 4.958 ± 1.268 and 4.857 ± 0.900, respectively, when compared to that of the inactive sub-class with corresponding value of 4.333 ± 0.577. Finally, analysis of ALogP revealed that the inactive sub-class afforded higher polarity with a mean of 2.459 ± 1.848 than both the intermediate and active sub-classes, which gave 3.634 ± 0.922 and 3.592 ± 0.537, respectively. A closer glance of the inactive sub-class shows that **32** afforded significantly higher ALogP value of 4.588 than the other two compounds (**26** and **33**) having values of 1.269 and 1.52.

### *Data pre-processing and feature selection*

Quantum chemical descriptors of compounds from low energy conformer of compounds derived from geometrical optimization at B3LYP/6-31g(d). Furthermore, low energy conformer was then subjected to further calculation using the Dragon software for generating an additional set of molecular descriptors. The usefulness of quantum chemical and molecular descriptors in constructing QSAR/QSPR models of chemical properties (Nantasenamat et al., 2005, 2007a, b) and biological activities (Khoshneviszadeh et al., 2012; Martinez-Martinez et al., 2012; Nantasenamat et al., 2010, 2013b; Uesawa et al., 2011; Worachartcheewan et al., 2011, 2013) had previously been demonstrated. Constant and multi-collinear descriptors were removed from the initially large set of 3,224 descriptors from Dragon resulting in a reduced subset of 1,129 molecular descriptors. Subsequently, this set of descriptors was then combined with a set of 6 quantum chemical descriptors to give rise to a total of 1,135 descriptors. Such combined set was then subjected to stepwise MLR for further round of feature selection thereby re-

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

sulting in 7 significant descriptors consisting of F10[N-O], Inflammat-50, Psychotic-80, H-047, BELe1, B10[C-O] and MAXDP. Tables 1 and 2 display the molecular descriptors constituting the data set and the definition of significant descriptors, respectively. An intercorrelation matrix of Pearson's correlation coefficient was constructed as to verify the independence of the constituent variables in this set of molecular descriptors. Results indicated that this set of 7 descriptors were independent from one another as deduced by the low correlation coefficient values (Table 3).

### *QSAR model of aromatase inhibitory activity*

The set of 7 significant descriptors were then used in the development of QSAR models using MLR to deduce a linear equation as described by Eq. (2). Thus, such multivariate analysis generated a total of 4 models consisting of MLR equations and statistical analysis as summarized in Table 4. Inherent outliers present in each of the MLR models were identified using absolute standardized residual of 2 as the cutoff.

The initial model 1 produced the following MLR equation and statistical parameters:

**Table 2:** Definition of important descriptors for QSAR model

| Symbol | Definition | Type |
|---|---|---|
| **F10[N-O]** | Frequency of N-O at topological distance 10 | 2D frequency fingerprints |
| **Inflammat-50** | Ghose-Viswanadhan-Wendoloski anti-inflammatory-like index at 50 % | Molecular properties |
| **Psychotic-80** | Ghose-Viswanadhan-Wendoloski anti-psychotic-like index at 50 % | Molecular properties |
| **H-047** | H attached to C1(sp3)/C0(sp2) | Atom-centred fragments |
| **BELe1** | Lowest eigenvalue n.1 of Burden matrix/weighted by atomic Sanderson electronegativities | Burden eigenvalues |
| **B10[C-O]** | Presence/absence of C-O at topological distance 10 | 2D binary fingerprints |
| **MAXDP** | Maximal electrotopological positive variation | Topological descriptors |

**Table 3:** Intercorrelation matrix of significant descriptors for QSAR model

| | F10[N-O] | Inflammat-50 | Psychotic-80 | H-047 | BELe1 | B10[C-O] | MAXDP |
|---|---|---|---|---|---|---|---|
| **F10[N-O]** | 1.0000 | | | | | | |
| **Inflammat-50** | -0.0636 | 1.0000 | | | | | |
| **Psychotic-80** | 0.0105 | -0.7346 | 1.0000 | | | | |
| **H-047** | 0.1205 | -0.5443 | 0.6005 | 1.0000 | | | |
| **BELe1** | 0.2980 | -0.1769 | 0.2311 | 0.5513 | 1.0000 | | |
| **B10[C-O]** | 0.0708 | 0.1166 | 0.2249 | 0.2614 | 0.1685 | 1.0000 | |
| **MAXDP** | 0.3587 | -0.1136 | 0.2099 | 0.4372 | 0.6917 | 0.3267 | 1.0000 |

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

$pIC_{50}$ = - 0.5797(F10[N-O]) + 1.0993(Inflammat-50) + 0.4107(Psychotic-80) + 0.1123(H-047) - 13.2991(BELe1) - 0.6687 (B10[C-O]) + 0.4336(MAXDP) + 29.0288 [3]

$N$ = 29, $R^2_{Tr}$ = 0.9036, $RMSE_{Tr}$ = 0.1412, $Q^2_{CV}$ = 0.7693, $RMSE_{CV}$ = 0.2238, $F$ ratio = 10.00, critical $F$ value = 2.488 and $R^2_{Tr}$ - $Q^2_{CV}$ = 0.1343

The model showed compounds **27** and **29** as outliers, which were removed from the model and the resulting data set was re-calculated to generate Model 2 as shown below.

$pIC_{50}$ = - 0.5407(F10[N-O]) + 1.0870(Inflammat-50) + 0.4025(Psychotic-80) + 0.1051(H-047) - 15.5871(BELe1) + 0.4285(MAXDP) + 32.8689 [4]

$N$ = 27, $R^2_{Tr}$ = 0.9347, $RMSE_{Tr}$ = 0.1179, $Q^2_{CV}$ = 0.8821, $RMSE_{CV}$ = 0.1585, $F$ ratio = 24.94, critical $F$ value = 2.599 and $R^2_{Tr}$ - $Q^2_{CV}$ = 0.0526

It was observed that B10[C-O] was removed from the MLR equation owing to the fact that the remaining 27 compounds had constant value of 1 for the B10[C-O] descriptor. Likewise, compound **20** was detected as the outlying compound and following its removal yielded model 3 as described below.

$pIC_{50}$= - 0.6008(F10[N-O]) + 1.0371( Inflammat-50) + 0.3941(Psychotic-80) + 0.1042(H-047) - 15.7039(BELe1) + 0.4959( MAXDP) + 32.7809 [5]

$N$ = 26, $R^2_{Tr}$ = 0.9496, $RMSE_{Tr}$ = 0.1082, $Q^2_{CV}$ = 0.9111, $RMSE_{CV}$ = 0.1423, $F$ ratio = 32.44, critical $F$ value = 2.628 and $R^2_{Tr}$ - $Q^2_{CV}$ = 0.0385

Subsequently, compound **18** was identified as the outlying compound and its removal resulted in model 4:

$pIC_{50}$ = - 0.6067(F10[N-O]) + 1.0108(Inflammat-50) + 0.4262(Psychotic-80) + 0.0884(H-047) - 14.5495(BELe1) + 0.4563(MAXDP) + 30.9432 [6]

$N$ = 25, $R^2_{Tr}$ = 0.9579, $RMSE_{Tr}$ = 0.0958, $Q^2_{CV}$ = 0.9239, $RMSE_{CV}$ = 0.1304, $F$ ratio = 36.42, critical $F$ value = 2.661 and $R^2_{Tr}$ - $Q^2_{CV}$ = 0.0340

This final model 4 was shown to afford robust predictive performance as verified from its statistical parameters for both the training and LOO-CV sets. Scatter plots of experimental versus predicted $pIC_{50}$ values for the 4 models of the training and LOO-CV sets are presented in Figures 3a-d as white and black squares, respectively. Interestingly, the statistical performance increased from Eq. 3 to Eq. 6 as deduced from the increasing $Q^2$ and $F$ ratio. A metric proposed by Eriksson and Johansson (1996) for evaluating the reliability of predictive models was obtained by calculating the difference of $R^2$ and $Q^2$. A value in excess of 0.3 is indicative of chance correlation or the presence of outliers. Thus, the $R^2$–$Q^2$ from the training and LOO-CV sets were found to be < 0.3 thereby indicating its statistical significance. The experimental and predicted $pIC_{50}$ values along with their respective residuals are provided in Table 5.

### *Validation of QSAR model using external set*

The predictivity of QSAR models 1-4 was verified by external validation using 15 % of samples from the original data set (i.e. comprising of compounds **4**, **6**, **12**, **17** and **26**). It should be noted that the external set was randomly selected and their distribution are shown in Figure 4 as black squares along with the internal set shown as white squares. The internal and external predictivity of models 1-4 were determined and results suggested good predictive performance especially by model 4, which provided $Q^2_{CV}$ = 0.9239, $RMSE_{CV}$ = 0.1304 and $Q^2_{Ext}$ = 0.7268. Model 4 was selected for further investigation owing to its absence of outlying compounds and its ability to perform well on both the internal and external sets. Moreover, $R^2$–$Q^2$ calculated from training and external sets afforded a value of < 0.3, which suggests the model's reliability.

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

**Table 4:** Summary of predictive performance of QSAR model

| Model[a] | Training set | | | LOO-CV set | | | External set | | |
|---|---|---|---|---|---|---|---|---|---|
| | *N* | $R^2_{Tr}$ | $RMSE_{Tr}$ | *N* | $Q^2_{CV}$ | $RMSE_{LOO-CV}$ | *N* | $Q^2_{Ext}$ | $RMSE_{Ext}$ |
| **1** | 29 | 0.9036 | 0.1412 | 29 | 0.7693 | 0.2238 | 5 | 0.4017 | 0.4514 |
| **2** | 27 | 0.9347 | 0.1179 | 27 | 0.8821 | 0.1585 | 5 | 0.7291 | 0.2986 |
| **3** | 26 | 0.9496 | 0.1082 | 26 | 0.9111 | 0.1423 | 5 | 0.7517 | 0.2815 |
| **4** | 25 | 0.9579 | 0.0958 | 25 | 0.9239 | 0.1304 | 5 | 0.7268 | 0.2927 |

[a] Compounds **18**, **20**, **27** and **29** as outliers were removed from the models.

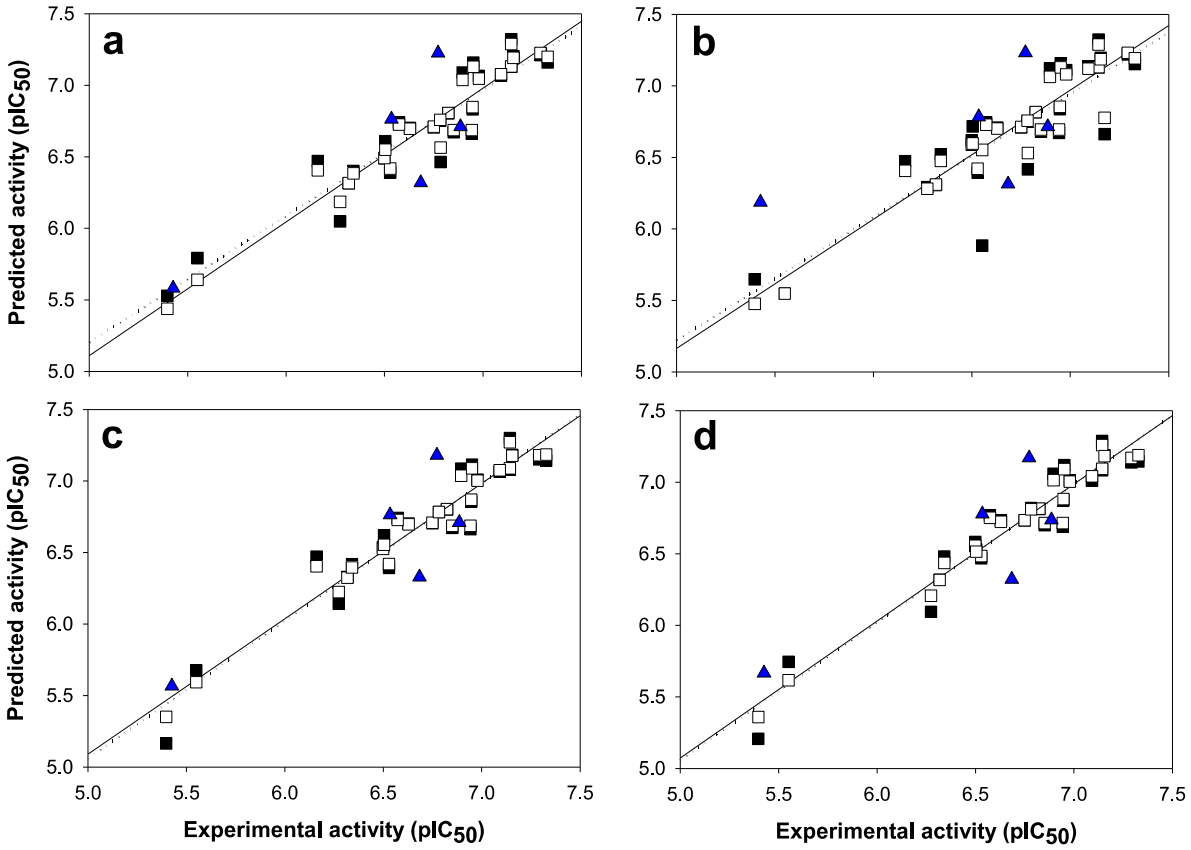



**Figure 3:** Plot of experimental versus predicted aromatase inhibitory ($pIC_{50}$) activity for the training set (white squares; regression line is represented as solid line), leave-one-out cross-validation set (black squares; regression line is represented as dotted line) and external test set (blue triangles) for model 1 (a), model 2 (b), model 3 (c) and model 4 (d)

Plots of experimental and predicted $pIC_{50}$ values for the 4 models of the external set are presented in Figures 3a-d as blue triangles. As previously mentioned for the internal set, the corresponding values of the experimental and predicted $pIC_{50}$ values and their respective residuals are shown in Table 5.

### *Elucidating structure-activity relationship*

Aromatase inhibitory potency ($IC_{50}$ value) of coumarin analogs was transformed to $pIC_{50}$ using Eq. (1) and was used as the dependent variable. Feature selection performed on a set of molecular and quantum chemical descriptors yielded 7 significant descriptors as follows (listed in order of rela-

tive importance as deduced from the MLR regression coefficient values): BELe1 (Burden eigenvalues) > Inflammat-50 (molecular properties) > B10[C-O] (2D binary fingerprints) > F10[N-O] (2D frequency fingerprints) > MAXDP (Topological descriptors) > Psychotic-80 (molecular properties) > H-047 (Atom-centered fragments).

Disubstituted coumarin pharmacophores bearing 7-ether linkage and 4-methylimidazole moieties displayed different potency in their aromatase inhibitory effect. The QSAR study revealed many significant descriptors plausibly governing its interaction with aromatase. Such descriptors provided information on the types of functional groups related to their lipophilicity, polarity and isomeric effects as well as appropriate distance of functional groups.

In the series of 7-benzyloxy courmarin (**1**-**13**, $R_1 = CH_2C_6H_5$, $R_2 = H$), BELe1 values weighted by atomic Sanderson electronegativities were shown to be in the range of 1.920-1.923 where the phenyl groups ($R_1$) were substituted with *m*– and *p*-Me, F, Cl, OMe, $OCF_3$ and $NO_2$. Interestingly, *m*-substituents such as F (**3**), Cl (**4**), $OCF_3$ (**6**) exhibited higher aromatase inhibitory activity than their corresponding *p*-substituent (**8**, **9**, **11**) counterparts. The *m*-F compound **3** was shown to be the strongest aromatase inhibitor but compound **11** bearing the *p*-$OCF_3$ moiety displayed the least activity. A closer look revealed that the compounds had the same value of BELel of 1.920 while affording different value for the maximal electrotopological positive variation (MAXDP) with corresponding values of 5.234, 4.937 and 4.914 for *m*-F (**3**), *m*-$OCF_3$ (**6**) and *m*-Cl (**4**), respectively. In comparison with the corresponding *p*-substituted phenyl compounds (**8**, **9**, **11**), their MAXDP values (4.947, 4.930, 4.910, respectively) were relatively less than those of the *m*-substituted compounds (**3**, **4**, **6**). It can thus be presumed that an unsymmetrical *m*-substituted phenyl ($R_1$) compounds (**3**, **4**, **6**) displayed higher MAXDP values as compared to a symmetrical *p*-substituted compounds (**8**, **9**, **11**). Similarly, the unsymmetrical 3′,4′-di F phenyl compound (**13**) had higher MAXDP value (5.271) than the symmetrical 3′,5′-di F analog (**12**) with MAXDP value of 4.870.

It was noted that for a series of 7-phenoxy coumarins (**14**-**24**, $R_1 = C_6H_5$, $R_2 = H$), the unsymmetrical 3′,4′-di F phenyl (**24**) had higher MAXDP value as compared to the symmetrical 3′,5′-di F analog (**23**). As for the series of $R_1$ ($CH_3$ and $C_6H_5$) containing chiral center at the C-4 position of the coumarin ring (**28**-**32**), the 3′,4′-di F phenyl analog **30** bearing phenyl group at the C-4 methine carbon atom had higher MAXDP value (5.503) when compared to the most potent aromatase inhibitor **24** (without chiral center) having MAXDP = 5.299. So far, the 3′,4′-di F phenyl analog **24** displayed 6.74 folds higher aromatase inhibitory activity than analog **30**. In addition, analog **24** exhibited higher activity (1.43 folds) than analog **29** containing phenyl moiety ($R^2$) at the C-4 methine carbon but without 3′,4′-di F substituted at the phenyl ring ($R_1$). The results indicated that 3′,4′-di F phenyl group ($R_1$) predominantly governed the aromatase inhibitory activity as compared to the phenyl ring ($R^2$). This could be attributed to the fact that compounds **29** and **30**, which contain steric effect from phenyl group at the C-4 methine carbon, could prevent its binding or interaction with the target site of action. The 7-phenoxy compound **24** (without phenyl group at the methine carbon) was reported to bind the aromatase active site (heme iron) via coordination with the lone pair electron from N atom of the imidazole ring in such a way that the imidazole ring at the C-4 position is perpendicular to the coumarin ring (Stefanachi et al., 2011).

Coumarin analogs ($R_1 = C_6H_5$) without the 3′,4′-di F group such as **31** ($R_2 = C_6H_4$-Cl-*p*) and **32** ($R_2 = C_6H_4$-CN-*p*) afforded high MAXDP (5.480 and 5.504) and high BELe1 (1.957 and 1.962) values, respectively. Notably, the inductive effect of the *p*-CN group (**32**) can provide the resonant ionic charge distribution form (**32a**, Figure 5) accounting for its higher MAXDP and BELe1

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

values than that of the *p*-Cl analog **31**. Amongst the 4,7-disubstituted coumarins (**1-32**), analog **32** was shown to be the least potent aromatase inhibitor in which such remarkably low activity of **32** might be due to the ionic charge **32a** and the bulky *p*-CN phenyl group (at chiral center).

On the other hand, 7-hydroxy **26** ($R_1$ = H) and methoxy **27** ($R_1 = CH_3$) coumarins showed low values of MAXDP (4.471 and 4.6000) and BELe1 (1.914 and 1.916). Such low values for these descriptors could be possibly a result from the lack of phenoxy group at the C-7 position (**26** and **27**) in contributing lone pair electron from the O atom into the phenyl ring ($R_1$) leading to higher topological positive variation as compared to that of the 7-phenoxy compound **14** that can provide positively-charged molecule **14a** (Figure 5).

Similarly, the electronic effect of the 7-phenylamino group (**25**) on the coumarin core structure could plausibly give rise to positively-charged species **25a** (Figure 5) with high MAXDP value of 4.890. It should be noted that all compounds (**1-34**) share the common ionic charge distribution form (**1A**) as shown in Figure 5.

However, the highest MAXDP value (5.972) was noted for the 3-substituted imidazolylcoumarin **34** as compared to the 4-substituted imidazolylcoumarin analog **29** (MAXDP = 5.447). This could probably be due to the distance or position of substituents on the coumarin ring that affected its resonant ionic charge formation. The 3-imidazole

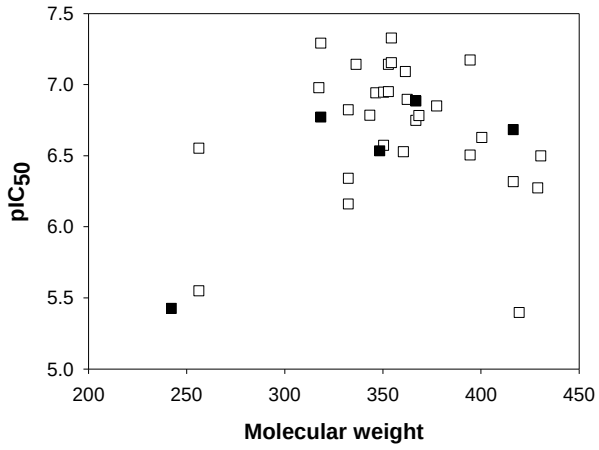



**Figure 4:** Plot of molecular weight versus aromatase inhibitory activity (pIC50) used in selection of external (black squares) and internal (white squares) sets

**Table 5:** Experimental and predicted aromatase inhibitory activities ($pIC_{50}$) of coumarin analogs (**1-34**) obtained from model 4

| Compound | Experimental $pIC_{50}$ | Predicted $pIC_{50}$ | Residual |
|---|---|---|---|
| **1** | 6.824 | 6.814 | -0.010 |
| **2** | 6.943 | 6.691 | -0.252 |
| **3** | 6.947 | 6.872 | -0.075 |
| **4**[a] | 6.886 | 6.737 | -0.149 |
| **5** | 6.629 | 6.734 | 0.105 |
| **6**[a] | 6.684 | 6.322 | -0.362 |
| **7** | 6.851 | 6.701 | -0.150 |
| **8** | 6.573 | 6.771 | 0.198 |
| **9** | 6.750 | 6.734 | -0.016 |
| **10** | 6.896 | 7.059 | 0.163 |
| **11** | 6.318 | 6.319 | 0.001 |
| **12**[a] | 6.772 | 7.170 | 0.398 |
| **13** | 6.783 | 6.82 | 0.037 |
| **14** | 7.292 | 7.139 | -0.153 |
| **15** | 7.143 | 7.084 | -0.059 |
| **16** | 7.143 | 7.289 | 0.146 |
| **17**[a] | 6.535 | 6.778 | 0.243 |
| **18** | 6.161 | ND | - |
| **19** | 6.951 | 7.120 | 0.169 |
| **20** | 6.785 | ND | - |
| **21** | 6.529 | 6.469 | -0.060 |
| **22** | 7.092 | 7.012 | -0.080 |
| **23** | 7.155 | 7.189 | 0.034 |
| **24** | 7.328 | 7.146 | -0.182 |
| **25** | 6.979 | 7.013 | 0.034 |
| **26**[a] | 5.426 | 5.666 | 0.240 |
| **27** | 6.553 | ND | - |
| **28** | 6.342 | 6.481 | 0.139 |
| **29** | 7.174 | ND | - |
| **30** | 6.499 | 6.583 | 0.084 |
| **31** | 6.274 | 6.095 | -0.179 |
| **32** | 5.397 | 5.207 | -0.190 |
| **33** | 5.550 | 5.744 | 0.194 |
| **34** | 6.504 | 6.530 | 0.026 |

ND = not determined owing to the fact that the compounds were outliers.
[a]Compounds were used as external set.

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

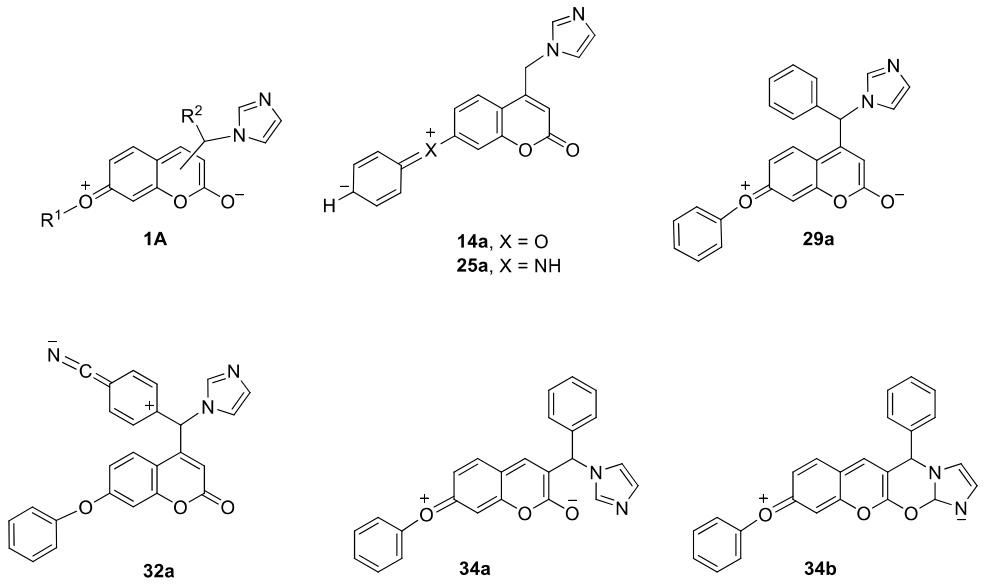



**Figure 5:** Chemical structures of ionically charged resonant forms

substituent of coumarin **34** lying in close proximity to the carbonyl lactone that induces higher charge distribution (**34a**) as deduced from the high MAXDP value as compared to the corresponding value of 4-substituted coumarin **29** showing ionic charge formation (**29a**) as illustrated in Figure 5. Apparently, the ionic form **34a** provided the oxy anion that could attack the imine moiety of the imidazole ring to form negatively-charged N atom on the imidazole moiety (**34b**, Figure 5). This may be accounted for by the higher MAXDP value of **34** than that of compound **29**.Taken together, the lipophilic property of $R_1$ substituents play crucial role in affording appropriately high values of MAXDP and BELe1 that accounted for the potent aromatase inhibitory effect. Particularly, the phenoxy analog **24** bearing 3′,4′-di F phenyl ($R_1$) exerted the most potent activity as compared to the corresponding 3′,4′-di F benzyl ($R_1$) analog **13**. This can be attributed to the correlation of descriptors as well as electronic effect of functional groups in contributing ionic charge resonant forms as well as the proper distance between $R_1$ and the coumarin core structure in which the phenoxy ($R_1$ = phenyl) coumarin analog gave the best fit in interacting with the target site.

### *Molecular docking of coumarins to aromatase*

Insights on the binding modality of coumarins to aromatase were elucidated by means of molecular docking. Particularly, low-energy conformers of coumarins obtained from aforementioned quantum chemical calculations were docked to the previously reported protein structure of aromatase (Suvannang et al., 2011) that had been subjected to 100 ns of molecular dynamics simulation. The azole moiety is central to the interaction of non-steroidal AIs with the iron atom of the metalloporphyrrins (Balding et al., 2008; Maurelli et al., 2011; Pearson et al., 2006). The handling of metal ions is not a trivial task and poses a great challenge in molecular docking and heme-containing proteins are particularly difficult to deal with owing to deficiencies in scoring function or the less stringency imposed by heme cofactors when compared to those of other metal ions (Irwin et al., 2005; Seebeck et al., 2008). Thus, to counter this limitation and ensure reliable interpretations we selected only docking conformations in which the az-

EXCLI Journal 2014;13:1259-1274 – ISSN 1611-2156
Received: July 08, 2014, accepted: November 18, 2014, published: December 08, 2014

ole moiety is oriented towards the heme co-factor for further analysis.

Binding poses with the lowest docked energy and belonging to the top-ranked cluster was selected as the final model for post-docking analysis with AutoDock Tools and PyMOL. Towards understanding the key residues involved in the protein-ligand interaction, we focused on residues within 4 Å around the docked ligands and this is depicted in Figure 6 where panel a shows all ligands docked to the aromatase binding pocket while panel b shows the most potent compound **24** as a representative example. Residues found interacting with most of the investigated coumarin chemotype were primarily lipophilic or non-polar residues comprising of Ile305, Ala306, Val370, Leu372, Met374, Leu477 and Ser478. Furthermore, polar and positively-charged residues included Thr310 and Arg115, respectively. Thus, it can be seen that aside from the azole-heme interaction and two polar residues, the protein-ligand interactions under investigation were predominantly lipophilic in nature. The aforementioned residues coincided with previously reported key residues in the binding cavity of aromatase (i.e. Ala306, Thr310, Met374 and Ser478) (Favia et al., 2013; Lo et al., 2013; Suvannang et al., 2011). Moreover, as the crystal structure suggests, the dominance of lipophilic residues could be accounted for by its role in forming an androgen-specific cleft that binds snugly to the androstenedione (Ghosh et al., 2009).

A closer analysis of the docking poses was carried out by stratifying this set of 34 coumarin analogs to 3 sub-classes (i.e. active, intermediate and inactive) on the basis of their $pIC_{50}$ in which compounds having values greater than 7 were considered active, less than 6 as inactive and values in between were designated as intermediate. The general key residues found in both active and inactive sub-classes consisted of Ala306, Thr310 and Met374. Particularly, active compounds were found to interact with Phe221, Trp224, Leu228, Ile305, Asp309, Leu477 and Ser478 however these residues were not involved in interacting with the inactives.

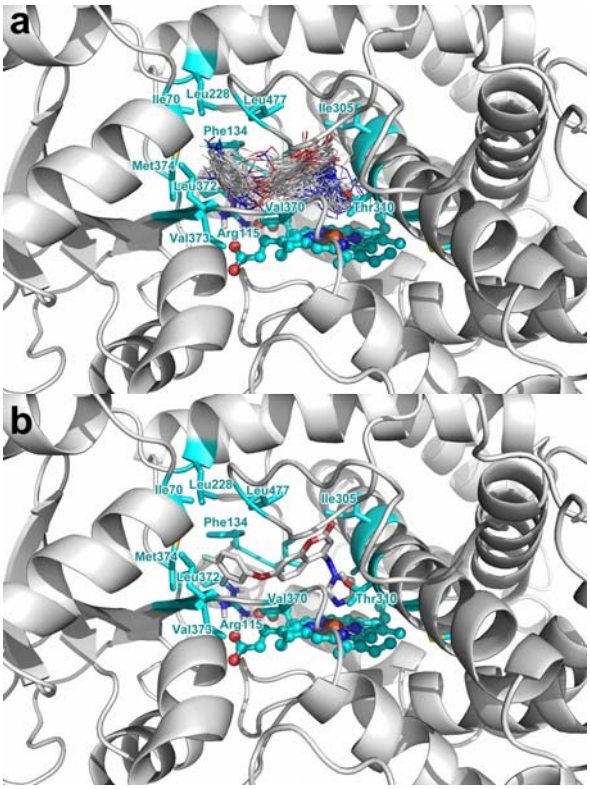



**Figure 6:** Docking poses of all 34 coumarins (a) and the most potent compound **24** (b) within the confinement of the aromatase binding pocket

## CONCLUSION

This study describes the QSAR study of imidazole derivatives of 3,7- and 4,7-disubstituted coumarins having $R_1$ and $R_2$ substituents as inhibitors of the aromatase enzyme. Significant molecular descriptors were identified to include F10[N-O], Inflammat-50, Psychotic-80, H-047, BELe1, B10[C-O] and MAXDP, which were used in the construction of QSAR models using the MLR method. Multivariate analysis afforded good predictive performance for the cross-validated internal set with $Q^2_{CV} = 0.9239$ and $RMSE_{CV} = 0.1304$ while an external validation confirmed its robustness with $Q^2_{Ext} = 0.7268$ and $RMSE_{Ext} = 0.2927$. Insights on the structure-activity relationship of compounds were also discussed in light of the selected set of significant descriptors in concomitant with structural details from substit-

uents $R_1$ and $R_2$. Both QSAR model and molecular docking investigations suggest that the aromatase inhibitory activity of compounds was primarily dependent on lipophilic properties and the position of substituent ($R_1$) on the coumarin core structure. Structural knowledge gained from QSAR models and molecular docking could be used to guide the rational design of novel aromatase inhibitors.

## ACKNOWLEDGEMENTS

This research project is supported by the Goal-Oriented Research Grant from Mahidol University to C.N. Talent Management fellowship to A.W. and research assistantship to N.S. are gratefully acknowledged.

## REFERENCES

Auclair K, Moenne-Loccoz P, Ortiz de Montellano PR. Roles of the proximal heme thiolate ligand in cytochrome p450(cam). J Am Chem Soc. 2001;123: 4877-85.

Balding PR, Porro CS, McLean KJ, Sutcliffe MJ, Marechal JD, Munro AW, et al. How do azoles inhibit cytochrome P450 enzymes? A density functional study. J Phys Chem A. 2008;112:12911-8.

Balunas MJ, Su B, Brueggemeier RW, Kinghorn AD. Natural products as aromatase inhibitors. Anticancer Agents Med Chem. 2008;8:646-82.

Bheemanapalli LN, Balakumar C, Kaki VR, Kaur R, Akkinepally RR. Pharmacophore based 3D-QSAR study of biphenyl derivatives as nonsteroidal aromatase inhibitors in JEG-3 cell lines. Med Chem. 2013; 9:974-84.

ChemAxon Ltd., MarvinSketch, Version 6.2.1. Budapest, Hungary, 2014. http://www.chemaxon.com.

Delano W. PyMOL, Version 0.99, DeLano Scientific LLC. Palo Alto, CA, 2002.

Desantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics. CA Cancer J Clin. 2011;61:409-18.

Eriksson L, Johansson E. Multivariate design and modeling in QSAR. Chemometr Intell Lab Syst. 1996;34:1-19.

Favia AD, Cavalli A, Masetti M, Carotti A, Recanatini M. Three-dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protoporphyrin IX for a molecular dynamics study of heme-cysteinato cytochromes. Proteins. 2006;62:1074-87.

Favia AD, Nicolotti O, Stefanachi A, Leonetti F, Carotti A. Computational methods for the design of potent aromatase inhibitors. Expert Opin Drug Discov. 2013;8:395-409.

Ferlin MG, Carta D, Bortolozzi R, Ghodsi R, Chimento A, Pezzi V, et al. Design, synthesis, and structure-activity relationships of azolylmethylpyrroloquinolines as nonsteroidal aromatase inhibitors. J Med Chem. 2013;56:7536-51.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09, Revision A.1. Wallingford, Connecticut, 2009.

Galeazzi R, Massaccesi L. Insight into the binding interactions of CYP450 aromatase inhibitors with their target enzyme:a combined molecular docking and molecular dynamics study. J Mol Model. 2012;18: 1153-66.

Ghosh D, Griswold J, Erman M, Pangborn W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. Nature. 2009;457:219-23.

Ghosh D, Lo J, Morton D, Valette D, Xi J, Griswold J, et al. Novel aromatase inhibitors by structure-guided design. J Med Chem. 2012;55:8464-76.

Guinez RF, Matos MJ, Vazquez-Rodriguez S, Santana L, Uriarte E, Olea-Azar C, et al. Synthesis and evaluation of antioxidant and trypanocidal properties of a selected series of coumarin derivatives. Future Med Chem. 2013;5:1911-22.

Hemshekhar M, Sunitha K, Thushara RM, Sebastin Santhosh M, Shanmuga Sundaram M. Kemparaju K et al. Antiarthritic and antiinflammatory propensity of 4-methylesculetin, a coumarin derivative. Biochimie. 2013;95:1326-35.

Irwin JJ, Raushel FM, Shoichet BK. Virtual screening against metalloenzymes for inhibitors and substrates. Biochemistry. 2005;44:12316-28.

Karelson M, Lobanov VS, Katritzky AR. Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev. 1996;96:1027-44.

Khoshneviszadeh M, Edraki N, Miri R, Foroumadi A, Hemmateenejad B. QSAR study of 4-aryl-4H-chromenes as a new series of apoptosis inducers using different chemometric tools. Chem Biol Drug Des. 2012;79:442-58.

Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA-a self-parameterizing force field. Proteins. 2002;47:393-402.

Krieger E, Nabuurs SB, Vriend G. Homology modeling. Methods Biochem Anal. 2003;44:509-23.

Lo J, Di Nardo G, Griswold J, Egbuta C, Jiang W, Gilardi G, et al. Structural basis for the functional roles of critical residues in human cytochrome p450 aromatase. Biochemistry. 2013;52:5821-9.

Martinez-Martinez FJ, Razo-Hernandez RS, Peraza-Campos AL, Villanueva-Garcia M, Sumaya-Martinez MT, Cano DJ, et al. Synthesis and in vitro antioxidant activity evaluation of 3-carboxycoumarin derivatives and QSAR study of their DPPH· radical scavenging activity. Molecules. 2012;17:14882-98.

Maurelli S, Chiesa M, Giamello E, Di Nardo G, Ferrero VE, Gilardi G, et al. Direct spectroscopic evidence for binding of anastrozole to the iron heme of human aromatase. Peering into the mechanism of aromatase inhibition. Chem Commun. 2011;47: 10737-9.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem. 1998;19: 1639-62.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785-91.

Nagamallu R, Kariyappa AK. Synthesis and biological evaluation of novel formyl-pyrazoles bearing coumarin moiety as potent antimicrobial and antioxidant agents. Bioorg Med Chem Lett. 2013;23:6406-9.

Nantasenamat C, Naenna T, Isarankura Na Ayudhya C, Prachayasittikul V. Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network. J Comput Aided Mol Des. 2005;19:509-24.

Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. Quantitative structure-imprinting factor relationship of molecularly imprinted polymers. Biosens Bioelectron. 2007a;22:3309-17.

Nantasenamat C, Isarankura-Na-Ayudhya C, Tansila N, Naenna T, Prachayasittikul V. Prediction of GFP spectral properties using artificial neural network. J Comput Chem. 2007b;28:1275-89.

Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Advances in computational methods to predict the biological activity of compounds. Expert Opin Drug Discov. 2010;5:633-54.

Nantasenamat C, Worachartcheewan A, Prachayasittikul S, Isarankura-Na-Ayudhya C, Prachayasittikul V. QSAR modeling of aromatase inhibitory activity of 1-substituted 1,2,3-triazole analogs of letrozole. Eur J Med Chem. 2013a;69:99-114.

Nantasenamat C, Li H, Mandi P, Worachartcheewan A, Monnor T, Isarankura-Na-Ayudhya C, et al. Exploring the chemical space of aromatase inhibitors. Mol Div. 2013b;17:661-77.

Narashimamurthy J, Rao AR, Sastry GN. Aromatase inhibitors: a new paradigm in breast cancer treatment. Curr Med Chem Anticancer Agents. 2004;4:523-34.

Narayana BL, Pran Kishore D, Balakumar C, Rao KV, Kaur R, Rao AR, et al. Molecular modeling evaluation of non-steroidal aromatase inhibitors. Chem Biol Drug Des. 2012;79:674-82.

Nativelle-Serpentini C, Moslemi S, Yous S, Park CH, Lesieur D, Sourdaine P, et al. Synthesis and evaluation of benzoxazolinonic imidazoles and derivatives as non-steroidal aromatase inhibitors. J Enzyme Inhib Med Chem. 2004;19:119-27.

Neves MA, Dinis TC, Colombo G, Sá e Melo ML. An efficient steroid pharmacophore-based strategy to identify new aromatase inhibitors. Eur J Med Chem. 2009;44:4121-7.

OpenEye Scientific Software, Babel, Version 3.3.0. Santa Fe, NM, 2014. http://www.eyesopen.com.

Pearson JT, Hill JJ, Swank J, Isoherranen N, Kunze KL, Atkins WM. Surface plasmon resonance analysis of antifungal azoles binding to CYP3A4 with kinetic resolution of multiple binding orientations. Biochemistry. 2006;45:6341-53.

Seebeck B, Reulecke I, Kamper A, Rarey M. Modeling of metal interaction geometries for protein-ligand docking. Proteins. 2008;71:1237-54.

Stefanachi A, Favia AD, Nicolotti O, Leonetti F, Pisani L, Catto M, et al. Design, synthesis, and biological evaluation of imidazolyl derivatives of 4,7-disubstituted coumarins as aromatase inhibitors selective over 17-alpha-hydroxylase/C17-20 lyase. J Med Chem. 2011;54:1613-25.

Sun M, Hu J, Song X, Wu D, Kong L, Sun Y, et al. Coumarin derivatives protect against ischemic brain injury in rats. Eur J Med Chem. 2013;67:39-53.

Suvannang N. Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Molecular docking of aromatase inhibitors. Molecules. 2011;16:3597-617.

Talete srl, DRAGON for Windows (Software for Molecular Descriptor Calculations), Version 5.5. Milano, Italy, 2007.

Thanikaivelan P, Subramanian V, Raghava Rao J, Unni Nair B. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. Chem Phys Lett. 2000;323:59-70.

Uesawa Y, Mohri K, Kawase M, Ishihara M, Sakagami H. Quantitative structure-activity relationship (QSAR) analysis of tumor-specificity of 1,2,3,4-tetrahydroisoquinoline derivatives. Anticancer Res. 2011;31:4231-8.

Varela C, Tavares da Silva EJ, Amaral C, Correia da Silva G, Baptista T, Alcaro S, et al. New structure-activity relationships of A- and D-ring modified steroidal aromatase inhibitors: design, synthesis, and biochemical evaluation. J Med Chem. 2012;55:3992-4002.

Venugopala KN, Rashmi V, Odhav B. Review on natural coumarin lead compounds for their pharmacological activity. Biomed Res Int. 2013;2013:963248.

Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques, 3rd ed. Burlington, MA: Morgan Kaufmann Publishers, 2011.

Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul S, Prachayasittikul V. Predicting the free radical scavenging activity of curcumin derivatives. Chemom Intell Lab Syst. 2011;109:207-16.

Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. QSAR study of amidino bis-benzimidazole derivatives as potent antimalarial agents against *Plasmodium falciparum*. Chem Pap. 2013;67:1462-73.

Worachartcheewan A, Nantasenamat C, Owasirikul W, Monnor T, Naruepantawart O, Janyapaisarn S, et al. Insights into antioxidant activity of 1-adamantylthiopyridine analogs using multiple linear regression. Eur J Med Chem. 2014;73:258-64.

Wu XQ, Huang C, Jia YM, Song BA, Li J, Liu XH. Novel coumarin-dihydropyrazole thio-ethanone derivatives: design, synthesis and anticancer activity. Eur J Med Chem. 2014;74:717-25.

Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. N Engl J Med. 2006;354:270-82.

Yoshioka S, Takahashi S, Hori H, Ishimori K, Morishima I. Proximal cysteine residue is essential for the enzymatic activities of cytochrome P450cam. Eur J Biochem. 2001;268:252-9.