

QSAR study of diarylalkylimidazole and diarylalkyltriazole aromatase inhibitors

Razieh Ghodsi^{1,2} · Bahram Hemmateenejad^{3,4}

Received: 27 June 2015 / Accepted: 8 February 2016 / Published online: 23 February 2016
© Springer Science+Business Media New York 2016

Abstract A QSAR analysis was conducted on a series of long-chained diarylalkylimidazole and diarylalkyltriazole derivatives as potent aromatase inhibitors. To obtain more appropriate QSAR models from a source of very large number of descriptors, a two-step stepwise variable selection strategy was performed. Firstly, from each group of the calculated descriptors, separate QSAR models were obtained. Then, the descriptors appeared in all of the generated models were subjected to another variable selection method and the obtained models were subjected to cross-validation. Finally, an external test set was used to access the ultimate performance of the models. The selected descriptors were analyzed for their influence on aromatase inhibition. The effects of hydration energy, position of H-bond acceptor, presence of cyano group, and shape of HOMO orbital on aromatase inhibition were successfully described, and they were consistent with the previous reports.

Keywords Aromatase · QSAR · Diarylalkylimidazole · Diarylalkyltriazole · Descriptor · HOMO

Electronic supplementary material The online version of this article (doi:10.1007/s00044-016-1530-1) contains supplementary material, which is available to authorized users.

✉ Bahram Hemmateenejad
hemmatb@sums.ac.ir

¹ Biotechnology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

² Department of Medicinal Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran

³ Chemistry Department, Shiraz University, Shiraz, Iran

⁴ Medicinal and Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

Introduction

The rate-limiting step in estrogen synthesis is the aromatization of androgen which happens due to aromatase activity. Aromatase enzyme converts C19 androgens to C18 estrogens and is responsible for the aromatization of androgens into estrogens by catalyzing three sequential hydroxylation reactions. Pathologically, an abnormal overexpression of aromatase in breast tissue is responsible for breast cancer development (Hong and Chen, 2006). Aromatase inhibitors are used clinically for the treatment of estrogen-dependent diseases such as breast cancer as an alternative to anti-hormone treatment (Brodie and Njar, 2000; Bayer *et al.*, 1991). The third-generation aromatase inhibitors, such as anastrozole, letrozole, and exemestane, are now considered as an alternative to tamoxifen in first-line treatment of advanced breast cancer (Wong and Ellis, 2004; Needleman and Tobias, 2008). So, increasing interest has been devoted to the synthesis of potent and selective inhibitors of aromatase by different research groups (Gobbi *et al.*, 2010; Doiron *et al.*, 2011; Ferlin *et al.*, 2013; McNulty *et al.*, 2014; Ghodsi *et al.*, 2016).

Nowadays, computational methods have found widespread application in drug design. They shorten the way of reaching to new drugs, and thus, they save in time and money. Among different computer-aided methods of drug design, quantitative structure–activity relationships (QSAR) are of special interests (Yousefinejad and Hemmateenejad, 2015). They are mathematical models relating biological activity to different structural aspects of the molecules. QSARs as promising tool in medicinal chemistry are used to predict the activity of new or even unsynthesized molecules and to describe the chemo-biological interactions involved in the biological activity under study.

2D-QSAR methods (Ghavami and Sepehri, 2016; Pasha *et al.*, 2008; Srivastava *et al.*, 2009) are robust and simple and do not need subjective molecular alignment or determination of 3D structures or putative binding conformation which is time-consuming. However, in some of 2D-QSAR methods such as Hansch–Fujita method, the overall predictive quality may not be as good as some multi-dimensional methods which are computationally more complex and demanding (Myint and Xie, 2010).

In the present article, we explore 2D-QSAR models for a series of long-chained diarylalkylimidazole and diarylalkyltriazole molecule skeletons, which were synthesized and evaluated for their ability to inhibit aromatase by Karjalainen *et al.* (2000). We selected these series because they were potent and also resembled potent aromatase inhibitors drugs such as letrozole and another point is that Karjalainen *et al.* also evaluated the anti-aromatase activity of some known potent aromatase inhibitors such as letrozole, fadrozole, CGS 18320B (compound 3), and compound 23 (Eli Lilly) and we used them in our study to know the requirements of the potent aromatase inhibitors.

Materials and methods

Data set

The biological data used in this article are 76 aromatase inhibitor molecules of long-chained diarylalkylimidazole and diarylalkyltriazole molecule skeletons (Karjalainen *et al.*, 2000). The basic structural skeleton of the compounds used in this study is given in Fig. 1. The detailed structural features and biological activity of these compounds are listed in Table S1 of the supplementary section. The biological activity data and the concentration of 50 % inhibition (IC_{50}) were converted to logarithmic scale (pIC_{50}) and then used for subsequent QSAR analysis as dependent variables.

Descriptor generation

The chemical structures of molecules were constructed using Hyperchem package (version 7, Hypercube Inc.) and were optimized with semi-empirical AM1 method. Then, different quantum chemical descriptors such as dipole moment (D.M) and energies of HOMO and LUMO orbitals, hydration energies, and lipophilicity (octanol/water partition coefficient; LogP) were calculated by the software. Dragon software was used for calculation of different 2D descriptors including functional groups, geometrical, topological, and constitutional descriptors for each molecule.

A total of 870 descriptors, categorized in 16 different groups, were calculated (see Table S2 of supplementary materials). In addition, the isosurface plots of HOMO energies around the atoms were constructed using Hyperchem.

Model development

QSAR models were developed using multiple linear regression (MLR) analysis coupled with variable selection method of stepwise selection. Since variable collinearity degrades the performance of MLR equations, the calculated descriptors were searched for collinearity, and among the detected collinear variables ($R^2 > 0.95$), the one represented the highest correlation with biological activity was retained and the rest were omitted. The data set was divided into training (61 molecules) and test (15 molecules) sets by sampling from the two-dimensional space of the principal component analysis scores of the descriptors (Figure S1 of supplementary section). We tried to sample the test set molecules from almost all parts of the shown distribution.

Since the number of the remaining descriptors was much higher than the number of the molecules in the training set, model development was achieved in a two-step manner Hemmateenejad and Yazdani, 2009). In the first step, separate QSAR models were developed from the pool of each group of descriptors. Then, the descriptors appeared in the QSAR models of different groups were used as a new source of descriptors to generate the final QSAR model. In the model development step, leave-one-out cross-validation (LOO-CV) was used to validate the obtained models to select the most convenient model. The final performances of the models were judged utilizing a separate test set.

Stepwise regression and principal component analysis were achieved using SPSS software (SPSS Inc., version 11). Some other calculations were done in MATLAB (Mathwork Inc., version 7) environment. A personal computer (CPU at 2.3 GHz) with Windows XP operating system was used throughout.

Results and discussion

Generally, the structure of nonsteroidal aromatase inhibitors consists of two parts: (1) the azole part with a nitrogen atom coordinating to the heme iron atom of aromatase and (2) the bulky aryl part, which is hydrophobic and acts like the same as the steroid ring of the substrate (Pouget *et al.*, 2004). In this study, the effect of different microscopic structural properties of the molecules including electronic,

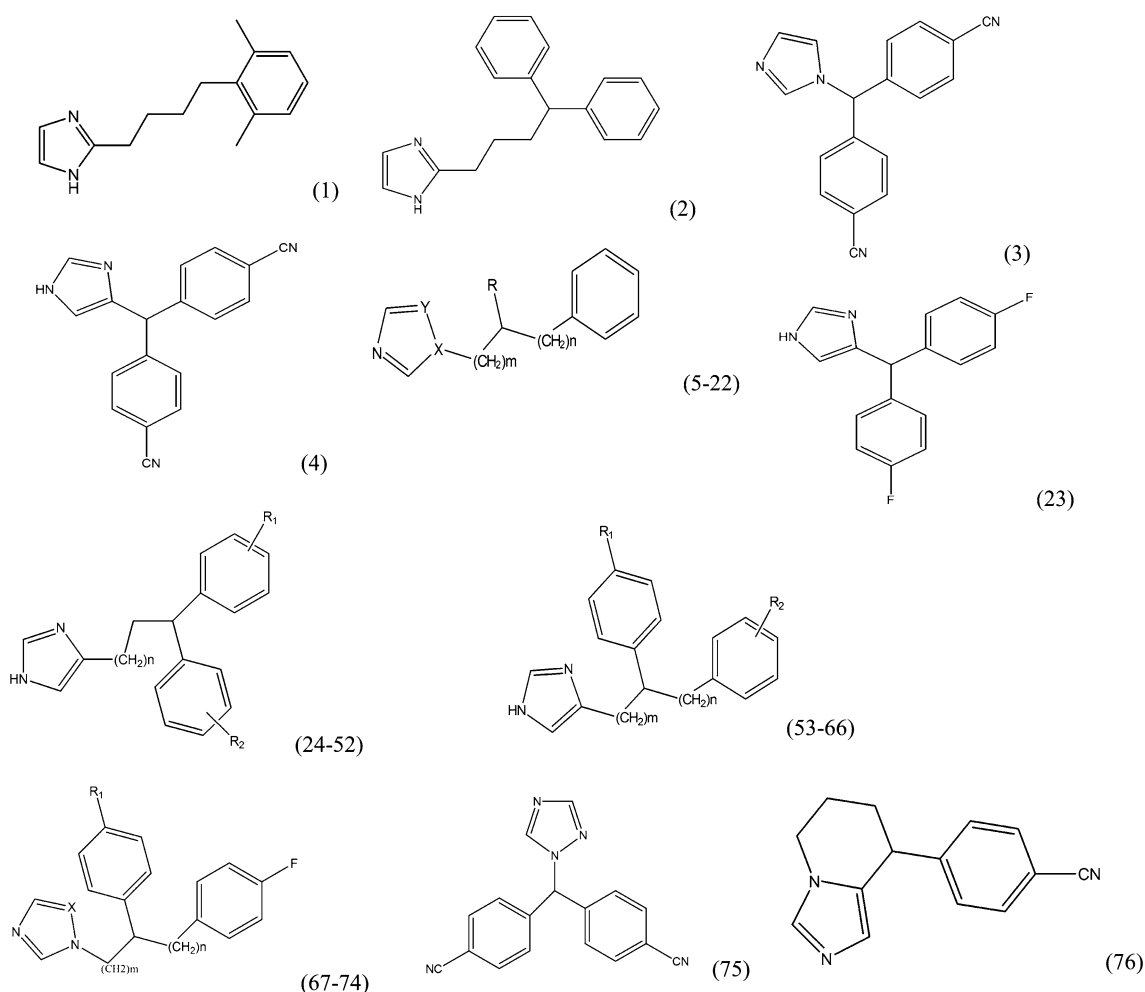


Fig. 1 Basic skeleton of the molecules used in this study

topologic, constitutional, and physicochemical descriptors on their aromatase activity has been investigated using a QSAR approach. The distribution of the studied molecules in the two-dimensional factor space of the calculated descriptors is given in Figure S1. It represents a relatively homogeneous distribution of molecules without any significant outlier. The test set molecules were sampled from this distribution.

The QSAR models obtained from the different groups of molecular descriptors are given in Table S2. The correlation coefficients of the models are in the range of 0.1–0.6. The number of selected descriptors is varied between 1 and 9. Models using higher number of variables resulted in better calibration statistics. However, models of higher calibration statistics do not mean that they are essentially better and more validation models should be developed. It should be noted that we would like to have as more useful variables as possible for the next step. Thus, we were not so serious in model validation and models of higher selected variables were considered for the future analyses.

In the previous step, a total number of 87 descriptors were identified as significant structural invariants affecting the aromatase inhibition of the studied molecules. In the next step, these descriptors were used to build the final QSAR model. Stepwise regression analysis suggested 11 significant QSAR models. To select the most convenient QSAR model, the models were validated by LOO-CV. Four models represented R_{CV}^2 of higher than 0.6, and thus, they were considered as predictive models according to the recommendation of Tropsha *et al.* (2003). These models in accompanying with their statistical parameters are given in Table 1. The reported P values suggest that all models are statistically significant with confidence level higher than 99.99 %. The calibration statistics (R^2 and SEC) are increased by increasing the number of variables. However, no significant changes are observed in the cross-validation statistics by increasing the number of variables from 8 to 11. To have a balance between the calibration and validation statistics and the number of variables, model number E9 with 9 input variables was selected for prediction of

Table 1 Final QSAR models and their statistical parameters

No.	Model	R^2	R^2CV	SEC	SECV	P
E8	$pIC_{50} = -13.62 + 1.57 \text{ nTB} + 4.48 \text{ BEHv2} + 2.00 \text{ ATS3m} - 0.72 \text{ ATS7m} - 0.39 \text{ nCp} - 0.05 \text{ G(N}\cdots\text{N)} - 0.10 \text{ Hyd.E} + 0.03 \text{ G(O}\cdots\text{O)}$	0.816	0.6729	0.3662	0.5166	0.001
E9	$IC_{50} = -20.34 + 1.48 \text{ nTB} + 4.915 \text{ BEHv2} + 2.00 \text{ ATS3m} - 0.71 \text{ ATS7m} - 0.39 \text{ nCp} - 0.05 \text{ G(N}\cdots\text{N)} - 0.10 \text{ Hyd.E} + 0.03 \text{ G(O}\cdots\text{O)} + 0.06 \text{ SRW07}$	0.8328	0.7047	0.3525	0.4768	0.001
E11	$IC_{50} = -21.52 + 1.45 \text{ nTB} + 4.62 \text{ BEHv2} + 2.08 \text{ ATS3m} - 0.75 \text{ ATS7m} - 0.39 \text{ nCp} - 0.05 \text{ G(N}\cdots\text{N)} - 0.10 \text{ Hyd.E} + 0.03 \text{ G(O}\cdots\text{O)} + 0.07 \text{ SRW07} + 0.56 \text{ ATS8v}$	0.8492	0.7356	0.3381	0.4605	0.001
E10	$IC_{50} = -15.61 + 1.512 \text{ nTB} + 3.69 \text{ BEHv2} + 2.00 \text{ ATS3m} - 0.71 \text{ ATS7m} - 0.43 \text{ nCp} - 0.05 \text{ G(N}\cdots\text{N)} - 0.09 \text{ Hyd.E} + 0.028 \text{ G(O}\cdots\text{O)} + 0.074 \text{ SRW07} + 0.571 \text{ ATS8v} - 14.665 \text{ G3p}$	0.8891	0.7071	0.3275	0.4904	0.001

new samples and for description of the studied interaction system.

The cross-validation predicted values of pIC_{50} for training molecules together with relative error of predictions are listed in Table S3. In addition, the prediction results for the external test set molecules, which did not contribute to the model development and selection, are given in Table S4. Obviously, for most of the test set molecules, the predictions are accurate and the prediction errors are in the range of the errors obtained for the training set. However, the prediction result for only one molecule (molecule 53) is associated with large error. The reported pIC_{50} for this molecule is 3.68, whereas the corresponding predicted value by QSAR model is 5.61 (52 % higher than experimental value). By comparing the IC_{50} of molecule 53 with its neighbors in Table 1, one can observe that the reported IC_{50} for molecule 53 is strangely much higher than the others. In Table 3 of the reference of this work (Karjalainen *et al.*, 2000), which the experimental data were collected, the IC_{50} of molecules 53 has been written 210 μM . However, according to the selectivity ratio of desmolase over aromatase for these molecules ($D/A = 9$) and the IC_{50} of this molecule toward desmolase, which is 19.0 μM , one can conclude that the IC_{50} of molecule 53 most probably is 2.1 μM and there is an error in the decimal point of IC_{50} of molecule 53 toward aromatase. If we accept this statement, the experimental pIC_{50} of molecule 53 would be 5.68, which is very close to the predicted value by the suggested QSAR model. So, for calculation of the statistical parameters of the prediction set, the pIC_{50} value of 5.68 was used. At last, to check our conclusion about this molecule, we found the original patent (Karjalainen *et al.*, 1995). In this patent, the IC_{50} of molecule 53 was reported to be 2.8 instead of 210, so the experimental pIC_{50} of molecule 53 is 5.55, which is also very close to the predicted value (5.61) by the suggested QSAR model (%E = + 1.06).

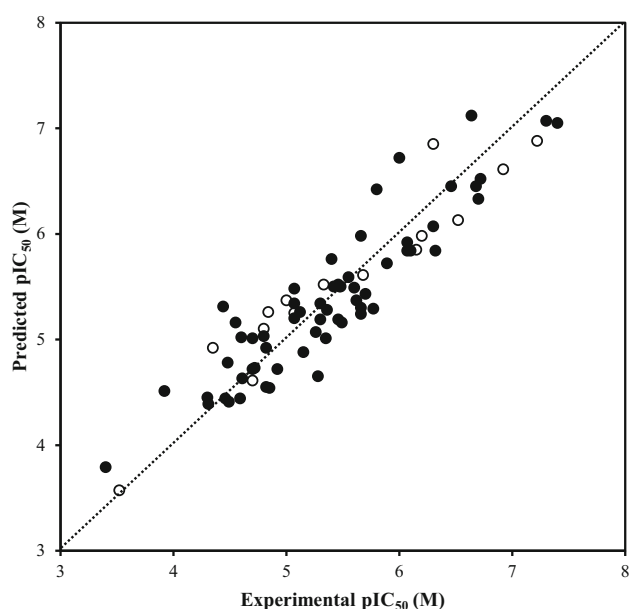
To have an overview on the overall prediction ability of the suggested QSAR model, it was validated according to

the criteria suggested by Tropsha *et al.* (2003), Golbraikh and Tropsha (2002), and Roy and Roy (2008). In particular, the correlation coefficient between the predicted and observed activities of compounds from an external test (r^2), the correlation coefficients for regressions through the origin (predicted versus observed activities or observed versus predicted activities, i.e., r_0^2 or $r_0'^2$, respectively), and the slope of the regression lines through the origin (k and k' , respectively) were calculated. Tropsha *et al.* (2003; Golbraikh and Tropsha, 2002) considered a QSAR model to be predictive, if all of the following conditions are satisfied: (i) $Q^2 > 0.5$, (ii) $r^2 > 0.6$, (iii) r_0^2 or $r_0'^2$ is close to r^2 , such that $[(r^2 - r_0^2)/r^2]$ or $[(r^2 - r_0'^2)/r^2] < 0.1$, and (iv) $0.85 < k < 1.15$ or $0.85 < k' < 1.15$. In addition, according to the recommendation of Roy and Roy (21), an additional statistic for external validation (rm^2) was calculated as $rm^2 = r^2 * [1 - (r^2 - r_0^2)^{1/2}]$. For a model with good external predictability, rm^2 value should be > 0.5 . The calculated parameters for the external test set molecules are given in Table 2. As seen, model 9 possesses all requirements to be accepted as a predictive model. Also, it is preferred for prediction ability over model 8. In addition, models of higher numbers of variables do not have much higher statistical parameters to be preferred over model 9. The plot of predicted activities against the experimental values for both training and test molecules is given in Fig. 2. Obviously, the data are distributed around the ideal straight line of zero intercept and unit slope. One can also observe a homogeneous distribution of training and test set data.

In addition to checking the prediction ability of the QSAR model 9, it was validated for chance correlation by y-randomization. The activity values were randomly shuffled 50 times, whereas the independent variables were kept constant. New regression models were generated using the randomly shuffled dependent variable, and then, correlation coefficient of prediction ($R_{P_rand}^2$) was calculated for each model. The average value of $R_{P_rand}^2$ for 50 randomized activity data was 0.273. The reliability of the

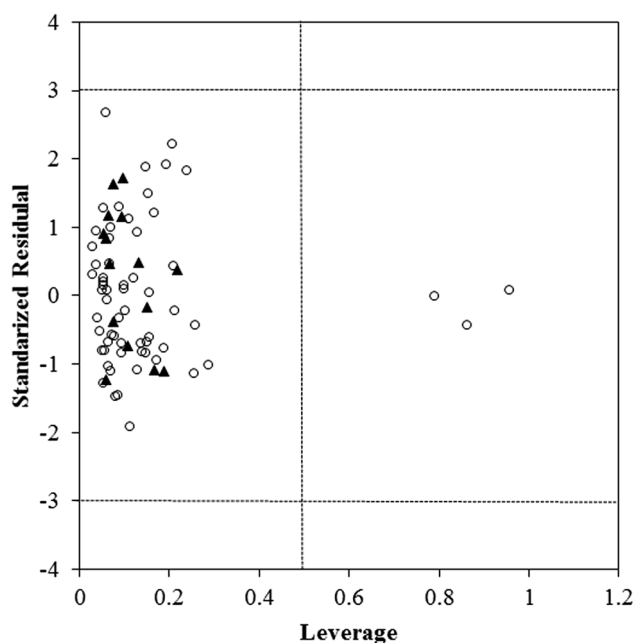
Table 2 Statistical parameters of the external test set

Parameter	Model			
	E8	E9	E10	E11
r^{2ab}	0.991	0.913	0.905	0.889
r_0^2	0.857	0.860	0.852	0.813
$r_0'^2$	0.898	0.900	0.894	0.874
$(r^2 - r_0^2)/r^2$	0.135	0.058	0.059	0.084
$(r^2 - r_0'^2)/r^2$	0.093	0.014	0.069	0.017
rm^2	0.628	0.702	0.850	0.640
rm'^2	0.689	0.808	0.840	0.770
k	0.987	1.007	1.005	1.006
k'	1.009	0.989	0.990	0.989

**Fig. 2** Plot of predicted activities by QSAR model E9 against the experimental value; *filled markers* are for cross-validation of training set, and *open markers* are for test set. The *dotted diagonal line* is an ideal fit with zero intercept and unit slope

QSAR model 9 (for not being obtained by chance) was confirmed using the corrected parameter (${}^cR_p^2$) (Mitra *et al.*, 2010). The minimum threshold value for this parameter is 0.5, and we obtained a ${}^cR_p^2$ value of 0.730.

The applicability domain (AD) of the QSAR model 9 was also investigated. The AD of a QSAR model is the range within which it tolerates a new molecule. There are different approaches for visualizing the AD (Roy *et al.*, 2015). Here, we utilized the most commonly used Williams plot (Garg and Smith, 2014) that is the plot of leverage against the standardized residual plot (Fig. 3). Leverage (h) shows how reliable is the prediction for new

**Fig. 3** Williams plot to evaluate the applicability domain of QSAR model 9 for training set (*open circles*) and test set (*closed triangles*). The *vertical dotted line* shows the warning leverage ($h^* = 0.49$), and the *horizontal dotted lines* show the range of accurate prediction within $\pm 3 U$

compounds, and hence, it is related to confidence interval of the predicted values. Prediction for compounds with h value lower than the warning leverage (h^*) is thought to be reliable. The warning leverage is defined as $3(p+1)/n$, where p and n are the number of descriptors and the number of molecules in the training set, respectively. In this essay, p and n are 9 and 61, respectively, and hence, h^* is calculated as 0.49. For the training set, compounds with $h > h^*$ are considered with largest structural influence in model development, whereas for the test set, the compounds $h > h^*$ are considered to be predicted unreliably.

On the other hand, residuals are related to accuracy and show how the predicted value is close to the actual value. The compounds with standardized residual no >3 U (\pm standard deviation) are considered to be predicted accurately. Figure 3 shows that the standardized residuals for both training and test compounds are within ± 3 U. This implies that QSAR model 9 predicted the activity of all compounds accurately, and thus, all compounds are inlier. Also, Fig. 3 suggests that all compounds in the data set are within the applicability domain except 3 molecules (#49, #50, and #76). These compounds have very low standardized residual, and hence, they can be considered as the influential compounds in the regression.

After approving the statistical significance and prediction quality of the suggested models, now it is time to investigate the role of the descriptors appeared in the QSAR models. A brief description of the descriptors used by QSAR models of Table 2 is given in Table 3. Here, we explain the contribution of some descriptors in the model. To support our obtained results, we used some cross-references.

The descriptor $G(N\cdots N)$ is sum of geometrical distances between N and N atoms, especially between the N of CN and imidazolic or triazolic N that coordinate with heme. The descriptor $G(O\cdots O)$ is defined as sum of geometrical distances between O and O atoms. These descriptors are important in aromatase inhibitory because oxygen of methoxy groups can be considered as the H-bond acceptor, depending on its spatial position relative to the nitrogen atom coordinating the iron ion. The methoxy groups can bind in different enzyme regions, being characterized by a different distance between the imidazolyl and the phenyl rings (Leonetti *et al.*, 2004).

Another factor in these equations is nCp (number of total primary C of sp³ hybridization). This suggests the importance of the length of the bridge carbon chain,

which represented negative effect on potency. The length of the bridge linking the imidazole or triazole to the phenyl rings is another important element affecting the enzyme affinity. In the other studies (Leonetti *et al.*, 2004; Recanatini *et al.*, 2001), the authors indicated that lacking the spacer results in a strong reduction in enzyme inhibition due to restriction of the conformational flexibility of compounds, whereas elongation of the spacer results in a reasonable decrease in inhibition. The length of the spacer can affect inhibitory potency by changing the position of the H-bond acceptors relative to imidazolic or triazolic N that coordinate with heme. It is known (Recanatini *et al.*, 2001) that the best spacer between imidazole or triazole and phenyl ring that had a H-bond acceptor in para position is one methylene, and in most of the nonsteroidal aromatase inhibitor drugs such as Arimidex, letrozole, and fadrozole, there is one methylene as a spacer (see the structure of letrozole and fadrozole in Fig. 1 and Table S1).

The descriptor nTB (number of triple bonds) is another influential parameter. Since the triple bond in the studied molecules is related to nitrile, nTB can be considered as the indication of the presence of CN, a strong H-bond acceptor, in molecular structure. Although the presence of a strong H-bond acceptor (like CN or F) is not strictly necessary to inhibit aromatase (Cavalli and Recanatini, 2002), they can increase the affinity of the compounds to aromatase by interacting with hydrophilic residue such as serine in the binding pocket of aromatase and as a result increase the inhibitory potency.

The negative sign of the coefficient of the hydration energy (Hyd. E) indicates that the potent molecules should have low values of the hydration energy. This means that lipophilicity can be one of the driving forces for the strong binding to aromatase. As Ghosh *et al.* (2010) revealed in the binding pocket of aromatase, the substrate is delineated

Table 3 Selected descriptors in final QSAR models

Descriptor	Definition
G3p	3rd component symmetry directional WHIM index/weighted by atomic polarizabilities
$G(N\cdots N)$	Sum of geometrical distances between N \cdots N
ATS8v	Boro-Moreau autocorrelation of a topological structure $-\text{lag}8/\text{weighted by atomic van der Waals volumes}$
SRW07	Self-returning walk count of order 07
$G(O\cdots O)$	Sum of geometrical distances between O \cdots O
ATS7m	Boro-Moreau autocorrelation of a topological structure $-\text{lag}7/\text{weighted by atomic masses}$
nCp	Number of total primary C(sp ³)
BEHv2	Highest eigenvalue n. 2 of burden matrix/weighted by atomic van der Waals volume
ATS3m	Boro-Moreau autocorrelation of a topological structure $-\text{lag}3/\text{weighted by atomic masses}$
nTB	Number of triple bonds
Hyd. E	Hydration energy

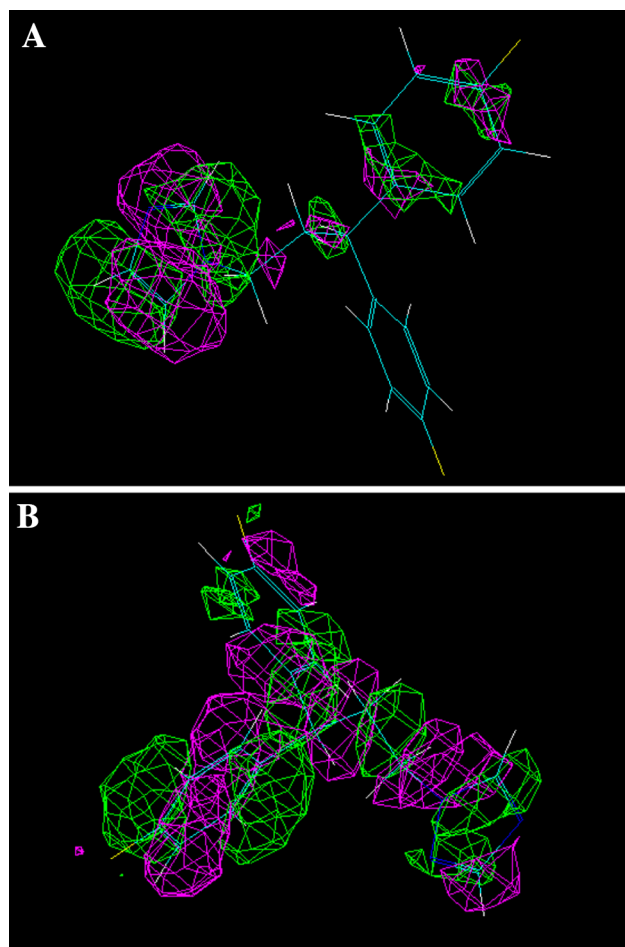


Fig. 4 HOMO shape of **a** molecule 68, imidazole derivative; $IC_{50} = 2.2 \mu\text{M}$ and **b** molecule 69, triazole derivative, $IC_{50} = 36 \mu\text{M}$

with a confluence of tight packing hydrophobic residues from different parts of the polypeptide chain.

Previous studies (Jones *et al.*, 1990; Okada *et al.*, 1996, 1997) revealed that inhibitory potency for aromatase depends on the position and number of nitrogen atoms in the heterocycles, which can be due to polar and charge interactions. We investigated the HOMO shape of imidazole derivative and corresponding triazole derivative (Fig. 3). It appears that the molecular orbital involved in the eventual coordination bond with the iron atom is located on the imidazole ring of molecule 68, but it is spread over the phenyl rings in the case of molecule 69. The relative unavailability of the nitrogen electron pair of the latter compound explains the low inhibitory activity of the 1,3,4-triazole-1-yl-substituted derivatives. However, there are several aromatase inhibitors drugs, possessing a triazole moiety, which are very potent, e.g., vorozole, letrozole. In fact, Lange *et al.* (1993) demonstrated that the replacement of the imidazole group by a triazole moiety results in lower in vitro inhibition and higher in vivo activity. It has shown that the increased in vivo activity of

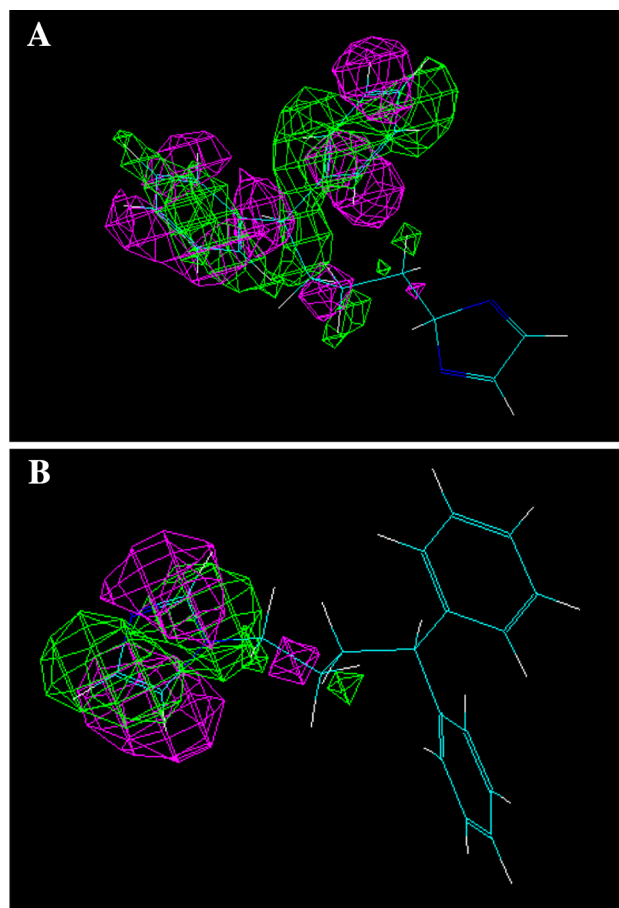


Fig. 5 Effect of 1- and 2 substitution on the HOMO shape of the two imidazole derivatives. **a** Molecule 22 with IC_{50} of $16 \mu\text{M}$ and **b** 2-(4,4-diphenylbutyl)-2H-imidazole with IC_{50} of $>1000 \mu\text{M}$

triazoles compared with imidazoles is due to increased metabolic stability of the triazole derivatives.

Karjalainen *et al.* found that the substitution of the 1-position of imidazole with the diarylalkyl moiety preserved some activity (e.g., see compound 22 with IC_{50} of $16 \mu\text{M}$), whereas the substitution of the 2-position of imidazole reduced activity radically (for example, 2-(4,4-diphenylbutyl)-2H-imidazole with IC_{50} over $1000 \mu\text{M}$). The difference in potency of the molecules can be explained partly by difference in the HOMO shape of the molecules (Fig. 4). In high-potency molecule, the HOMO orbital is mainly located over the imidazolyl ring, whereas in the low potency molecule, the HOMO orbital is distributed over other parts of the molecules (Fig. 5).

Conclusion

A QSAR study was conducted on the series of aromatase inhibitors in a two-step manner. The final models were validated rigorously, and among the generated models, a

9-parametric model was selected as the most appropriate model of high goodness of fit and external prediction. The suggested QSAR model could extensively describe the factors affecting the aromatase inhibitory activity of the studied molecules. The importance of sum of geometrical distances between two oxygen or two nitrogen atoms in the molecules was related to hydrogen bonding interaction of the inhibitors with aromatase in a special geometry. The presence of CN, a strong H-bond acceptor, in molecular structure was another influential parameter. Also, the number of bridging methylene groups in the space was found as another significant parameter. Our study indicates lipophilicity can be one of the effective factors for the strong binding to aromatase. Finally, the shape of HOMO orbital and its distribution over imidazole ring were described as the other important factors. The results obtained in this study were consistent with previous works. On the other hand, we used the cross-references to support our findings and conclusion in this study.

References

- Bayer H, Batzl C, Hartmann RW, Mannschreck A (1991) New aromatase inhibitors. Synthesis and biological activity of pyridyl-substituted tetralone derivatives. *J Med Chem* 34:2685–2691
- Brodie AM, Njar VC (2000) Aromatase inhibitors and their application in breast cancer treatment. *Steroids* 65(4):171–179
- Cavalli A, Recanatini M (2002) Looking for selectivity among cytochrome P450s inhibitors. *J Med Chem* 45:251–254
- Doiron J, Soultan AH, Richard R, Touré MM, Picot N, Richard R, Čuperlović-Culf M, Robichaud GA, Touaibia M (2011) Synthesis and structure–activity relationship of 1- and 2-substituted-1,2,3-triazole letrozole-based analogues as aromatase inhibitors. *Eur J Med Chem* 46:4010–4024
- Ferlin MG, Carta D, Bortolozzi B, Ghodsi R, Chimento A, Pezzi V, Moro S, Hanke N, Hartmann RW, Basso G, Viola G (2013) Design, synthesis, and structure–activity relationships of azolylmethylpyrroloquinolines as nonsteroidal aromatase inhibitors. *J Med Chem* 56:7536–7551
- Garg R, Smith CJ (2014) Predicting the bioconcentration factor of highly hydrophobic organic chemicals. *Food Chem Toxicol* 69:252–259
- Ghavami R, Sepehri B (2016) QSPR/QSAR solely based on molecular surface electrostatic potentials for benzenoid hydrocarbons. *J Iran Chem Soc* 13:519–529
- Ghodsi R, Azizi E, Ferlin MG, Pezzi V, Zarghi A (2016) Design, synthesis and biological evaluation of 4-(Imidazolylmethyl)-2-aryl-quinoline derivatives as aromatase inhibitors and anti-breast cancer agents. *Lett Drug Des Discov* 13:89–97
- Ghosh D, Griswold J, Erman M, Pangborn W (2010) X-ray structure of human aromatase reveals an androgen-specific active site. *Steroid Biochem Mol Biol* 118:197–202
- Gobbi S, Zimmer C, Belluti F, Rampa A, Hartmann RW, Recanatini M, Bisi A (2010) Novel highly potent and selective nonsteroidal aromatase inhibitors: synthesis, biological evaluation and structure–activity relationships investigation. *J Med Chem* 53:5347–5351
- Golbraikh A, Tropsha A (2002) Beware of q₂. *J Mol Graph Model* 20:269–276
- Hemmateenejad B, Yazdani M (2009) QSAR models for half-wave reduction potential of steroids: a comparative study between feature selection and feature extraction from subsets of or entire set of descriptors. *Anal Chim Acta* 634:27–35
- Hong Y, Chen S (2006) Aromatase inhibitors: structural features and biochemical characterization. *Ann N Y Acad Sci* 1089:237–251
- Jones CD, Winter MA, Hirsch KS, Stamm N, Talor HM, Holden HE, Davenport JD, Krumkalns EV, Suhr RG (1990) Estrogen synthetase inhibitors. 2. Comparison of the in vitro aromatase inhibitory activity for a variety of nitrogen heterocycles substituted with diarylmethane or diarylmethanol groups. *J Med Chem* 33:416–429
- Karjalainen A, Pelkonen O, Sodervall ML, Lahde MA, Lammintausta RAS, Karjalainen AL, Kalapudas AM (1995) Aromatase inhibiting 4(5)-imidazoles. US Patent 5,439,928
- Karjalainen A, Kalapudas A, Sodervall M, Pelkonen O, Lammintausta R (2000) Synthesis of new potent and selective aromatase inhibitors based on long-chained diarylalkylimidazole and diarylalkyltriazole molecule skeletons. *Eur J Pharm Sci* 11:109–131
- Lang M, Batzl C, Furet P, Bowman R, Hausler A, Bhatnagar AS (1993) Structure–activity relationships and binding model of novel aromatase inhibitors. *J Steroid Biochem Mol Biol* 44:421–428
- Leonetti F, Favia A, Rao A, Aliano R, Paluszczak A, Hartmann RW, Carotti A (2004) Design and synthesis, and 3D QSAR of novel potent and selective aromatase inhibitors. *J Med Chem* 47:6792–6803
- McNulty J, Keskar K, Crankshaw DJ, Holloway A (2014) Discovery of a new class of cinnamyl-triazole as potent and selective inhibitors of aromatase (cytochrome P450 19A1). *Bioorg Med Chem Lett* 24:4586–4589
- Mitra I, Saha A, Roy K (2010) Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol Simul* 36:1067–1079
- Myint KZ, Xie XQ (2010) Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *Int J Mol Sci* 11:3846–3866
- Needleman SJ, Tobias JS (2008) Aromatase inhibitors in early hormone receptor-positive breast cancer: what is the optimal initiation time for the maximum benefit? *Drugs* 68:1–15
- Okada M, Yoden T, Kawaminami E, Shimada Y, Kudoh M, Isomura Y, Shikama H, Fujikura T (1996) Studies on aromatase inhibitors. I. Synthesis and biological evaluation of 4-amino-4H-1,2,4-triazole derivatives. *Chem Pharm Bull* 44:1871–1879
- Okada M, Yoden T, Kawaminami E, Shimada Y, Kudoh M, Isomura Y (1997) Studies on aromatase inhibitors. II. Synthesis and biological evaluation of 1-amino-1H-1,2,4-triazole derivatives. *Chem Pharm Bull* 45:333–337
- Pasha FA, Muddassar M, Beg Y, Cho SJ (2008) DFT-based de novo QSAR of phenoloxidase inhibitors. *Chem Biol Drug Des* 71:483–493
- Pouget C, Yahiaoui S, Fagnere C, Habrioux G, Chulia AJ (2004) Synthesis and biological evaluation of 4-imidazolylflavans as nonsteroidal aromatase inhibitors. *J Bioorg Chem* 32:494–503
- Recanatini M, Bisi A, Cavalli A, Belluti F, Gobbi S, Rampa A, Valenti P, Palzer M, Paluszczak A, Hartmann RW (2001) A new class of nonsteroidal aromatase inhibitors: design and synthesis of chromone and xanthone derivatives and inhibition of the P450 enzymes aromatase and 17 alpha-hydroxylase/C17, 20-lyase. *J Med Chem* 44:672–680
- Roy P, Roy K (2008) On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 27:302–313
- Roy K, Kara S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst* 145:22–29

- Srivastava HK, Pasha FA, Mishra SK, Singh PP (2009) Novel applications of atomic softness and QSAR study of testosterone derivatives. *Med Chem Res* 18:455–466
- Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSAR models. *QSAR Comb Sci* 22:69–77
- Wong ZW, Ellis MJ (2004) First-line endocrine treatment of breast cancer. *Br J Cancer* 90:20–25
- Yousefinejad S, Hemmateenejad B (2015) Chemometrics tools in QSAR/QSPR studies: a historical perspective. *Chemom Intell Lab Syst* 149:177–204